OF

ER

DOCUMENT RESUME

ED 187 883 CE 025 409

AUTHOR	Botterbusch, Karl F.
TITLE	Psychological Testing in Vocational Evaluation.
INSTITUTION	Wisconsin Univ. - Stout, Menomonie. Dept. of Rehabilitation and Manpower Services. Materials Development Center.
SPONS AGENCY	Rehabilitation Services Administration (DHEW), Washington, D.C.
PUB DATE	Dec 78
GRANT	12-P-55307/5
NOTE	101p.
AVAILABLE FROM	Materials Development Center, Stout Vocational Rehabilitation Center, University of Wisconsin--Stout, Menomonie, WI 54741 ($2.00)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Diagnostic Tests; Evaluation; *Occupational Tests; *Psychological Testing; *Standardized Tests; Testing; Test Norms; Test Reliability; Test Reviews; *Test Selection; Test Validity; Verbal Tests; *Vocational Rehabilitation
IDENTIFIERS	*Vocational Evaluation

ABSTRACT

This publication is designed to help vocational evaluators wisely select and use tests within the context of the referral questions and the individualized evaluation plan. The first of two parts contains information on why tests are used in evaluation, problems with tests, and how to select tests. Part 2 contains a review of specific tests that either have been found to be successful within vocational evaluation or have this potential. A total of thirty-six tests are included in the following areas: achievement batteries and reading tests; character and personality; intelligence; multi-aptitude batteries; vocations: clerical; vocations: interests; vocations: manual dexterity; and vocations: mechanical ability. For each test, the following information is provided: purpose, final score, description (including administration, content, and scoring), materials required, appropriate groups, technical considerations (including norm groups, reliability, and validity), American Psychological Association level, sources of information, availability, and comments. The addresses of the test publishers are appended. (LRA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187883

Psychological Testing

in

Vocational Evaluation

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Fry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Karl F. Botterbusch, Ph.D.

CE 025 409

ADDITIONAL COPIES OF THIS PUBLICATION MAY BE
PURCHASED FROM MDC FOR $2.00 PER COPY

INTRODUCTION

A review of the vocational evaluation literature in the area of psychological testing indicates that evaluators have what can only be described as a love-hate relationship with testing. Many articles denounce testing, use it as a scapegoat for client and program failure and, most significantly, claim that reaction to the testing movement is what really started vocational evaluation's fascination with work sample techniques. Yet at the same time evidence indicates that tests are widely used in vocational evaluation (Pruitt, 1972; Botterbusch, 1974) and that evaluators feel that they need to know more about the selection, administration, and interpretation of tests (Egerman and Gilbert, 1969; Sankovsky, 1971; Task Force No. 3, 1975; Coffey, 1978).

This apparent conflict can be resolved by a careful reassessment of the role of tests in vocational evaluation. Tests per se are not "good" or "bad," "useful" or "not useful." Granted that some tests either by careful design and/or by continuing research are more widely and successfully used than others, it is the improper use of the test by the evaluator, teacher, psychologist, etc., that causes the problem and not the test itself. Or to paraphrase a slogan of the National Rifle Association (the other NRA), tests don't fail people, people fail people. A test is a neutral object to be used or misused. The professional test user must ask in what situation and with what particular person is there a good chance that a certain test will yield accurate and useful information. Asking this specific question of a test (or any assessment technique) will result in a more useable answer than an emotionally laden question about the use of tests in general.

The test must be used in a planned situation. There must be reasons for testing as well as for using any other evaluation technique. These reasons should be based upon the individual evaluation plan as defined by the Commission on the Accreditation of Rehabilitation Facilities (CARF, 1978):

> 3.4.3.1.1.3 Based on referral information, the initial interview, and determined objectives, a specific written evaluation plan for each individual shall be developed. This plan shall:
>
> a. identify the questions to be answered through the evaluation.
>
> b. indicate how these questions will be answered.
>
> c. where appropriate, specify persons (staff, family, etc.) who will be involved in carrying out the plan. There should be evidence that these individuals are aware of their role in carrying out this plan.
>
> d. be periodically reviewed and modified as necessary. (p. 28)

The evaluator has at his disposal tests, work samples, situational assessment, and job site evaluation. The selection of general techniques and individual assessment tools for obtaining answers to specific evaluation questions eliminates: (1) the arbitrary administration of a battery of tests or work samples, (2) the systematic procedure of having every client go through the same string of experiences, and (3) an evaluation plan that does not show the client how this experience relates to his overall rehabilitation program.

The purpose of this publication is to help the evaluator wisely select and use tests within the context of the referral questions and the individualized evaluation plan. Part I will contain information on why tests are used in evaluation, problems with tests, and how to select tests. Part II is a careful review of specific tests that either have been found to be successful within vocational evaluation or have this potential. Also included is a list of books on testing and statistics. Part II will use the same format, and may be considered as a revision of the 1973 MDC publication, Tests and Measurements for Vocational Evaluators.

Finally, I would like to acknowledge the work done by two graduate assistants, Ms. Sharon Vasholz and Ms. Mary K. Gorine who located tests, prepared references, and helped in many other ways.

Karl F. Botterbusch, Ph.D.
December, 1978

CONTENTS

Page

Introduction

Part I

Why are Tests Used in Evaluation? 1

Problems With Test Use in Evaluation. 2

How to Select Tests 5

Specific Uses of Tests in the Evaluation Process. 7

A Model for Test Use. 9

Part II

Test Review Outline 13

Reviews of Tests

Achievement Batteries and Reading Tests

Adult Basic Learning Examination. 15
California Achievement Tests. 17
Gray Oral Reading Test. 20
Nelson-Denny Reading Test 22
Peabody Individual Achievement Test 24
Tests of Adult Basic Education. 26

Character and Personality

Edwards Personality Inventory 28
Edwards Personal Preference Schedule. 31
Minnesota Multiphasic Personality Inventory 33
Sixteen Personality Factor Questionnaire. 35
Work Environment Preference Schedule. 37

Intelligence

Culture Fair Intelligence Test. 39
Peabody Picture Vocabulary Test 41
Revised Beta Examination. 43
SRA Pictorial Reasoning Test. 45
SRA Verbal Form 47

Multi-Aptitude Batteries

Differential Aptitude Test. 49
USES General Aptitude Test Battery. 51
USES Nonreading Aptitude Test Battery 54

Page

Vocations - Clerical

General Clerical Test 56
Minnesota Clerical Test 58
SRA Typing Skills 60
Stenographic Aptitude Test. 62

Vocations - Interests

AAMD-Becker Reading-Free Vocational Interest Inventory. 64
Kuder Occupational Interest Survey. 66
Minnesota Importance Questionnaire. 68
Ohio Vocational Interest Survey 70
Strong-Campbell Interest Inventory. 72
Wide Range Interest-Opinion Test. 74

Vocations - Manual Dexterity

Crawford Small Parts Dexterity Test 76
Hand-Tool Dexterity Test. 78
Purdue Pegboard 80
Stromberg Dexterity Test. 82

Vocations - Mechanical Ability

Bennett Mechanical Comprehension Test 84
Revised Minnesota Paper Form Board Test 86
SRA Mechanical Aptitudes. 88

Sources of Information About Testing. 90

Addresses of Test Publishers. 95

References. 96

PART I

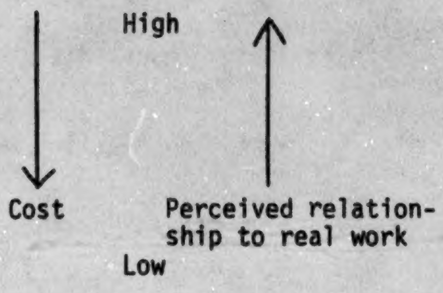
Why Are Tests Used in Evaluation?

Although this question has been commonly asked by many authors, to some extent it is not relevant, or in light of the introduction, should be rephrased to state: Is a particular test useful in certain vocational evaluation situations? While the merits of any specific test must be judged within the unique situation of its use, the concept of "tests" and their usefulness in vocational evaluation may be described more generally. So why use tests? There are several reasons.

The first is economic--compared with other evaluation techniques, tests (assuming that the evaluator can get useful information from a particular test) are the cheapest way of obtaining information about a client. Job site evaluation requires a heavy staff investment in terms of the amount of time needed to develop the initial job site and to maintain a client on the site. Situational assessment requires the existence of a workshop, production contracts, and staff for supervision. Both require that the client be paid for what he produces. If "homemade" work sample techniques are used, they require staff time to develop as well as money and time to construct; commercial work sample systems can cost up to $20,000 for the initial purchase. All work samples eventually require replacement parts and many require expendable supplies. Tests have the advantages of: (1) being cheaper to buy, (2) usually being group administered, and (3) often having separate answer sheets which reduce the expense of expendable supplies. In discussing these four evaluation techniques, there is probably an inverse relationship between the closeness of the technique to real work as viewed by the client and the overall cost of the technique. This can be visualized as follows:

Technique

Job Site Evaluation
Situational Assessment
Work Sample Techniques
Psychological Testing



There is a second economic factor--speed of administration. Most tests are desgined to take a sample of a behavior (or behaviors) in a relatively short period of time. While other assessment techniques may obtain more accurate information or more detailed information, they usually take longer than a test does. An evaluator can assess mechanical comprehension using a test that takes 30 minutes to administer to seven clients or he can place these clients on a work sample that assesses mechanical comprehension in two hours for each client. If only economical considerations were taken into account, then most evaluators would use tests as their only technique, thus becoming psychometricians. However, while no one can argue with the economic advantages of tests, there is a more basic consideration--are the results accurate for a particular client? If the evaluator wants to assess the mechanical comprehension of a client who can't read, who has a high degree of test anxiety, or who doesn't understand the instructions, then the most accurate method of assessment is not the cheapest or the shortest. Given the recommendations that are made from evaluation results, it may not be the most economical to the client or the rehabilitation services system in the long run. The point is that while tests are economical, they are only economical if the client can be accurately assessed by this method. In other words, if a test does not provide us with accurate information about a particular client, it is useless regardless of how economical it is to administer and score.

There are some situations where tests may be the most practical method of obtaining information. For example, the use of tests to determine the degree of literacy is important if the client is being considered for a formal training program or general academic work. Some covert personality characteristics that might not be readily inferred from behavioral observations may be brought to the attention of the evaluator through the use of tests. This could play an important role in determining what work adjustment services are needed.

Another reason for testing is to compare the client's performance with the most appropriate norm group, such as a group reading at the tenth grade level or workers employed in competitive industry. (The selection of the test and, subsequently, the norm group(s) depends in part on the decisions that are to be made from the test results.) If the goal of rehabilitation is to place the client in competitive employment, then all vocational evaluation techniques should compare the client's results with those employed in competitive work. We need to know how the client measures up against people who are competitively employed, who are in job training programs, etc. Careful selection of tests with appropriate norms is one way for the evaluator to get a good idea how his client compares with various non-disabled groups (assuming the evaluator selects a test with adequate norms and uses it properly). Such a comparison results in a more realistic evaluation of the client's potential.

One comment must be made about testing and the time period spent in vocational evaluation. It could be argued that one advantage of tests is that they can assess a person in a short period of time. With the reality of the one or two week evaluations in some facilities, this sounds like a good reason. However, tests should not be used only because they require less time than other techniques. Rather, a specific test should be given when the results are the best method of answering a referral question. With respect to the time to complete an evaluation, the length of time should be based on the client's needs and not on the calendar's.

> 3.4.3.1.1.4 The length of time an individual remains in vocational evaluation shall be primarily based upon the time necessary to accomplish the individual's evaluation goals. (CARF, 1978, p. 28)

Problems With Test Use in Evaluation

The first section pointed out some of the reasons why testing can be useful to vocational evaluators as well as mentioning that there are serious problems associated with test use. Generally, the failure of a test to yield accurate results may be based on two problems: (1) poor test usage and (2) some tests are not too useful for anything. This section deals with some of the ways tests can be misused. The next section deals with methods of selecting tests so that the poor ones can be avoided. Some of the common misuses of tests in vocational evaluation are:

Overuse ("Tuesday everyone takes the WRAT.") - Some facilities overuse tests by scheduling every client to take a certain test or group of tests. This overuse occurs for several reasons: (1) lack of individualized client planning, (2) a desire for easier client scheduling, and (3) faith that one test or specific battery of tests can yield valid and useful results for every client. There should be a reason why a test is given to a client just the same as there should be a reason why every other technique is used. To give a client, who the evaluator knows from prior information is illiterate, an achievement battery is a waste of time and money and only gives the client one more bad experience with tests. To be more general, before using any evaluation technique, see what information is needed for the individual client. To give an intelligence test to a client who has recently taken the WAIS is duplicating efforts. While it is true that some tests have more uses than others, it does not follow that a particular test is useful for every client. In conclusion, the best way to prevent overuse is to carefully plan each client's evaluation program.

Indiscriminate Use ("She was given a battery of clerical tests even though she had no interest in this area and the results of clerical work samples had been negative.") - Tests can be indiscriminately used as well as overused. The indiscriminate use of tests simply means that tests are used without much, if any, planning. In some cases a client may be administered a test or group of tests even when other evaluation results have already given the answer. If a client has taken several work samples in a specific area and if results indicate no interest and/or low performance, then there would appear to be little or no need to test the client in the same area. The converse is also true. Testing would not be needed if the client has shown previous interest and ability in one job area. For example, an automobile mechanic with a recent injury to his dominant hand is referred to find out if he can return to his former job. The first phase of this client's evaluation plan would center on the question of "Does he have the physical ability to return to his former occupation of mechanic?" This phase could be answered by work samples and job site evaluation. If the client could return to his former job, then no further assessment would be needed--certainly no tests. However, if this client cannot return to his previous job, then the second phase would be to answer the question "Is his ability and experience applicable to any jobs that are less physically demanding?" During the second phase, interest inventories and tests of mechanical and reasoning ability could prove useful. To give these or any other measures during the first phase would not be a wise use of testing. From this rather simple example two conclusions can be drawn: (1) testing should be carefully planned for each client and (2) testing need not be used when the evaluator is certain that other techniques have already given accurate information.

Separation of Learning From Performance ("Tests measure what the client does; not what he knows.") - A problem with testing, as with any other evaluation technique, is to separate learning from performance. This is sometimes difficult because learning cannot be seen, heard, or evaluated--only performance can be measured. Many problems in test use happen because the test administrator is not able to separate learning and performance. In the context of this publication, learning means that the evaluator communicates to the client what the test instructions are and how the client is to go about taking the test. Performance means the answers that the client gives to the test items. To have a client take a test before he is aware of what is involved, to have a client not understand the instructions, to have a client miss items because he cannot perceive them, to have a client "fail" a test because the format of the items was beyond his comprehension is to invalidate test results. Before the evaluator can measure the client's performance on a test, he must make sure that the client knows how to respond to the test items and that he accurately perceives the test content. In short, the client's test score must not reflect how well he understood the instructions or could perceive the items, but how well the client knew the answers "once he fully understood the instructions and the item content." If the purpose of the test is to obtain accurate information about the client, then the evaluator must make sure that the client knows the instructions and how to respond to the test content per se. It is the responsibility of the evaluator to make sure that the client knows what is required of him. This can be done in several ways: (1) careful test selection to pick tests with appropriate reading levels, etc., (2) the increased use of practice exercises and examples prior to actual administration, (3) the use of tests specially designed for certain disability groups, and (4) the modification of tests for certain disability groups. Test selection will be dealt with in this publication. The other three topics have been dealt with in two other MDC publications: The Use of Psychological Tests With Individuals Who Are Severely Disabled (Botterbusch, 1976) and Work Sample Evaluation of Blind Clients: Criteria for Administration and Development (Dickson, 1976). In short, one of the major reasons tests yield inaccurate information is because the test per se or the evaluator does not allow for the client to know what is expected and required of him prior to taking the actual test.

Negative Connotations of Tests by Evaluators and Clients ("Our facility doesn't like to use tests" and "I don't like to take tests.") - To many clients, as well as evaluators, the mere mention of the word "test" brings about negative connotations often accompanied by stronger emotions such as anxiety. Unfortunately many of these feelings are justified. All too often clients have had a history of failure in school tests, not being selected for jobs due to tests, and being classified almost automatically into a "failure" category. On the other side, some evaluators have often seen negative results from using tests, have little knowledge of test selection and use, and have serious doubts about the usefulness of tests. If tests are to be used successfully by the facility, the evaluator will have to deal with his own problems as well as those of the client. The evaluator can overcome his own problems in three ways: (1) by knowing how to select specific tests to suit specific client needs; (2) by testing only when necessary and then having the test and reason for its use written into the evaluation plan; and (3) by becoming more generally aware of the proper uses of testing. The evaluator can help the client overcome any mistrust of tests only after the evaluator himself is able to look at testing with an honest eye. After this occurs, the evaluator who chooses to use tests can help the client in several ways. First, provide the client with the exact reason for using each test. This reason is essentially what should have already been written on the individualized client evaluation plan. Second, for some tests, such as aptitude and achievement tests, provide extra occupational information. Tell the client in simple language what a sales aptitude test measures and how the results relate to a variety of sales jobs. Third, prior to the test tell the client how the results will be used and after the test tell the client exactly what the results are and what they mean. If these recommendations are followed, the client should respond positively and, most importantly, with increased motivation.

Poor Selection of Tests ("I'll see what we have around here to get an IQ score.") - Test use begins with selection and selection is based upon two questions: What do we want to find out about the client? and secondly, What tests will answer this? In answering the first question, the evaluator must be able to clearly identify what information about the client is needed to answer questions. In relating this to test use, the evaluator can begin by giving the general specifications where a test can be useful. For example, a picture interest inventory may be needed to help low literacy clients clarify their interests. After deciding what is needed and why it is needed, the evaluator searches for the best test that meets his specifications. In this example he will consider the range of jobs covered, the organization of the results, the ease of scoring as well as the technical questions of reliability, validity, and norms. Often several different tests are needed to answer the same questions for different client groups. Returning to the above example, if a general measure of client interest is needed, then the evaluator will want to have a wide range of instruments available--some for low literacy clients, for persons expressing a desire for a technical or skilled occupation, and for persons who want additional academic training. Thus, the measure of interest should consider what client is being assessed and how this assessment should take place. A quote from an article by Botterbusch and Sax (1977) dealing with the selection of commercial vocational evaluation systems also applies to test selection and use within the vocational evaluation process:

> The first decision is whether the evaluation unit is meeting client needs in terms of accurate assessment for available jobs and/or training. If needs are not being met, the second decision becomes a question of what areas of job assessment are needed for the evaluation unit. After these needs are known a thorough review of . . . available resources is made to determine how to best meet these needs . . . There has to be a great deal of analysis of what is to be accomplished during evaluation, the available jobs and training, the types of clients with whom you are working and the best way to accomplish the goals of your facility. This analysis is absolutely necessary before you can . . . [select any test]. (pages 35-36)

In conclusion, the evaluator must carefully select tests prior to their use. This selection should be based on what tests will provide answers to the referral questions. This section has presented some of the general problems with test use. The next section will deal in greater detail with how to select tests.

How to Select Tests

Test selection is not simple and many factors (ranging from technical considerations, through reading levels, to should tests be used at all) must be weighed. To be an "expert" in test selection, the evaluator must combine a detailed knowledge of two areas--technical knowledge about psychometrics and a practical knowledge of how to plan and conduct evaluation services for clients with a wide range of handicaps. During this review and selection process, the evaluator must keep in mind that the procedure must be based on what will best assess the client. While it is beyond the scope of this publication to present all the information needed to make any person an expert in test selection (although some source materials given on pages 90 to 94 will help you get started), some general considerations are as follows:

Should Tests Be Used ("Maybe some people shouldn't be tested at all.") - The first question to be asked when selecting a test (or tests) for a specific client population is not "what tests to use," but "should we use tests at all?" A review of the literature on test use shows that they can be adapted and/or designed for severely disabled client groups such as blind (Bauman, 1976; Scholl and Schur, 1976), deaf (Levine, 1976), and mentally retarded. While tests can be used with many severely disabled persons, the question is still to be asked, is a test usable with a particular client? In trying to answer this question, the evaluator must ask what potentially prevents the use of any tests with this client. A client may, for example, suffer from anxiety so intense that testing would be invalid except for all but the simplest motor tasks. A severely mentally retarded person from a culturally deprived background could perhaps be assessed more accurately using other techniques. The same is true for a person who is both blind and deaf. These are only examples. The point is that the evaluator must be able to discriminate those clients for whom testing of any kind would not produce accurate results. Therefore, the first step in test selection is to ask what tests would give accurate information of a particular quality for a particular client. If the evaluator finds that no test will provide this information, then he should use other assessment techniques.

Ask Questions About the Test ("What does this test really do?") - In selecting a test for potential use, the evaluator must carefully review the test, its manual, answer sheet, and other parts. The reviewer should look at the reading level, administration procedures, clarity of instructions, norms, reliability, and validity. The evaluator could begin by reading the stated purpose of the test in the manual, then reviewing the test itself, plus any research to determine if the test achieves its stated purpose. Critical reviews of the test in the Mental Measurements Yearbooks and elsewhere should be read. An obvious point is to consider lightly the publishers' advertising. When reviewing a new test for possible use, the purpose for the test within the evaluation unit and the general characteristics of the client population should be kept in mind. In other words, the test should be evaluated on its technical merits and on the place it will fill in the evaluation unit. (Part II of this publication is designed to provide some of the answers to these questions.)

Test Taking Ability ("Can the client take this test?") - The first two suggestions asked if the test should be used and what the test measured. This suggestion moves away from the test and centers on the client. Here we ask if the client can take the test in a way that yields accurate results. The first question is, can the client read the items and understand what he is reading? Most of the commonly used tests require the client to read (and comprehend) at the sixth or eighth grade level. If a client can understand the instructions (as evidenced by the practice items), can read the items with comprehension, and can use the appropriate answering format, then he can take most standardized tests. The client who has this test taking ability most likely will not need any low literate tests or special modifications in the instructions or answering method. This represents a great advantage to both client and evaluator. If the evaluator decides to use some tests with an individual client, then the first step is to determine the client's ability to take tests. Here the evaluator has several methods. The most straightforward is to use an established reading test or achievement battery to determine reading and comprehension, compare his score with the appropriate group, and make a decision. Subjective observations made during this initial test will give some indication of the client's test sophistication, the presence of any unusual degree of test anxiety, and the general ability to cope in a highly structured setting. Other methods of determining literacy may include checking previous records for recent testing, job history, and years of formal education. A client having a history of employment in clerical, sales, or technical occupations will probably read fairly well. Formal education may be another good estimate if it is recent. However, because of the social promotion policies of many schools, some high school graudates can't read at the sixth grade level. Test taking ability needs to be measured first. Once the literacy level is established, then the evaluator can plan any necessary testing around the client's level of reading and test sophistication.

Specific Uses of Tests in the Evaluation Process

The previous sections have mentioned that tests may often be useful at various steps of the evaluation process. This section will attempt to give some specific examples, but remember that these are only examples. Sometimes tests must be used to establish the legal "existence" of mental retardation for a school system, vocational rehabilitation agency, etc. If the evaluator is also qualified to administer the WAIS (Weschler Adult Intelligence Scale), WISC (Weschler Intelligence Scale of Children), or other individually administered instruments, then this testing may be done in the evaluation unit. If not, a psychometrician working closely with the unit is a good second alternative. The test results can be used for other aspects of the evaluation or to develop the evaluation plan. The second use of tests is to help plan the evaluation. If the referral contains little in the way of direct questions, then the evaluator could use testing at the start of the evaluation process to narrow down the client's interests, aptitudes, and to determine his literacy skills. After interviewing the client and assessing his literacy skills, a multi-aptitude battery together with a broad based interest inventory may be important. The results of these must be discussed with the client and used to help plan the rest of the evaluation. A note of caution is necessary--starting a new client with a battery of tests can easily fall into the problems of overuse and indiscriminate use. Thus, tests should be used with caution and also interspersed with other evaluation and information giving (e.g., occupational information) programs.

Tests can also be used to substantiate other evaluation results. If a client has done poorly on several work or job samples involving finger and hand dexterity, the evaluator may want to further investigate this by using a finger dexterity measure to determine the precise problems and to compare the client's performance with norms developed on employed workers. A client having done poorly on work samples in several occupational areas whose only common element may be the ability to visualize three-dimensional objects from diagrams and blueprints, could be given a test of spatial ability to see if this is the reason for the poor results. These two examples have used tests to check on the possible reasons why a client has experienced low achievement on another assessment technique. It should also be noted that other assessment techniques can be used to substantiate tests. Thus, we are not setting up tests or any other technique as the ultimate criteria. The process is to check the results of several measures against each other for consistency and, if inconsistent, to try to understand why.

Testing can also be used to help decide between alternatives. A client does well on a variety of work samples in both the clerical and sewing areas and has verbally expressed an interest in both. Other results agree that both areas are within the client's overall ability, range of dexterity, etc. In helping the client decide between these two areas, several methods may be used. Additional occupational information is given, the local job market is discussed and the chances for upward mobility are mentioned. At this point a more covert determination of interests and needs can be established using testing. Here tests could provide useful information not elsewhere attainable.

All these sources of information can be weighed by the evaluator and client to make a decision. Deciding between alternatives can also involve deciding between levels of jobs within an occupational (i.e., career) hierarchy. A client may be definitely interested in machine shop work but is somewhat confused as to the level. Jobs in this area vary in the degree of skill required and in the ability needed to reach this skill. Tests of mechanical comprehension, practical measurement of arithmetic, and ability to make fine discriminations can help the client decide if he should consider a fairly low level job such as a punch press operator, a semiskilled job like a drill press operator, or train for a position as a tool and die maker.

A final example is the use of tests to help answer questions or hypotheses raised by the client's history or evaluation results. A client has a job history of several clerical jobs held for short durations. Clerical tests and work samples reveal no skill deficiencies, behavioral observations show no overt problems that cause the client to loose a job, and the client is unable to adequately explain the reasons for the employment history. The following hypotheses could be set up: (1) the client has a personality or psychiatric problem that only manifests itself while on the job; (2) the client's intelligence may be extremely high causing boredom resulting in frequent job changes; and (3) the client may not really be interested in clerical work. The use of tests to answer these hypotheses would be discussed with the client and then one hypothesis would be investigated at a time. For the first hypothesis, a test like the 16PF or the MMPI could be useful; any of a number of intelligence measures could be used for the second; and the Strong, OVIS, or MIQ would help answer the third. The point is that the evaluator who encounters problems which appear to have no obvious solutions, could use testing as a method of seeing if his "hunches" are correct.

The above paragraphs have given some examples of the specific uses of testing during the evaluation process. These are only examples and the evaluator will be able to think of other situations where testing may be useful to help solve many other difficult questions. The themes that hopefully come through in this section are that tests can be used in creative and versatile ways and that if used correctly, a test can be an important part of many clients' evaluation plans.

A Model for Test Use

Throughout these pages we have talked about many factors affecting the use of tests in vocational evaluation. These can be divided into (1) the use of tests within an evaluation program and (2) the use of a specific test that will serve the needs of a specific client. Successful testing is administering the right test to the client so that accurate and useful information can be obtained. The left side of Figure 1 (i.e., "The Test") shows that tests are part of the total evaluation program and that the evaluation program must refer to the client population served, the local labor market (including educational opportunities), and the implicit or explicit philosophy of the evaluation unit. In establishing (or refining) an evaluation program, Step 1 is to decide if some of the information needed to assess clients may be obtained through testing. If the evaluator believes that some tests will be helpful to his clients, then the next step (Step 2) is to set up specifications for what sorts of tests are needed. Some examples of these specifications would be: (1) a test of mechanical aptitude that does not require any reading skills; (2) a test of literacy for low functioning persons that does not have childish content; and (3) a nonverbal test of general learning ability that has Spanish language directions. These examples are general and the evaluator may want to include other specifications such as cost, scoring procedures, cultural bias, and validity.

After setting up specifications, the evaluator finds and reviews tests that fit these specifications (Step 3). This publication will help the evaluator find some of these tests. Another way is to write the numerous test publishers, state what is needed, and ask for a catalogue. (The addresses of some test publishers are found on page 95.) Other methods of locating potentially useful tests include various books on testing, professional publications, and, especially, the Mental Measurements Yearbooks. After tests that may meet the specifications have been identified, the evaluator orders specimen sets from the publisher and then carefully reviews all of the material. An interesting procedure is to take the test yourself. This will help you to see the content of each item, to estimate its reading level, and to judge the appropriateness of the content for specific handicapped population(s). It may also be possible to try out the test on some clients; however, if this is done, make it very clear to the clients that the test is being evaluated and not them. After taking the test, score it and then interpret the results. These subjective methods of reviewing a test must be combined with a careful and critical review of the manual and materials. Some of the questions to be asked are about the stated purpose of the test, the technical aspects, reading level, administration and scoring time, type of administration, and interpretation of results. In reviewing any test the potential user may want to prepare a written outline of the critical points; the test review outline on page 13 can serve as a guide.

After reviewing the proposed test, the evaluator must decide if this test will be useful in obtaining accurate information for some of his clients (Step 4). This decision is based on a critical review of the test materials, the needs of the evaluation program, and the evaluator's knowledge of the characteristics of the clients who will be taking the test. If the decision is to reject a specific test, then other potential tests are reviewed. If there are no other potential tests, the evaluator may choose either another assessment technique or revise specifications for selecting a test.

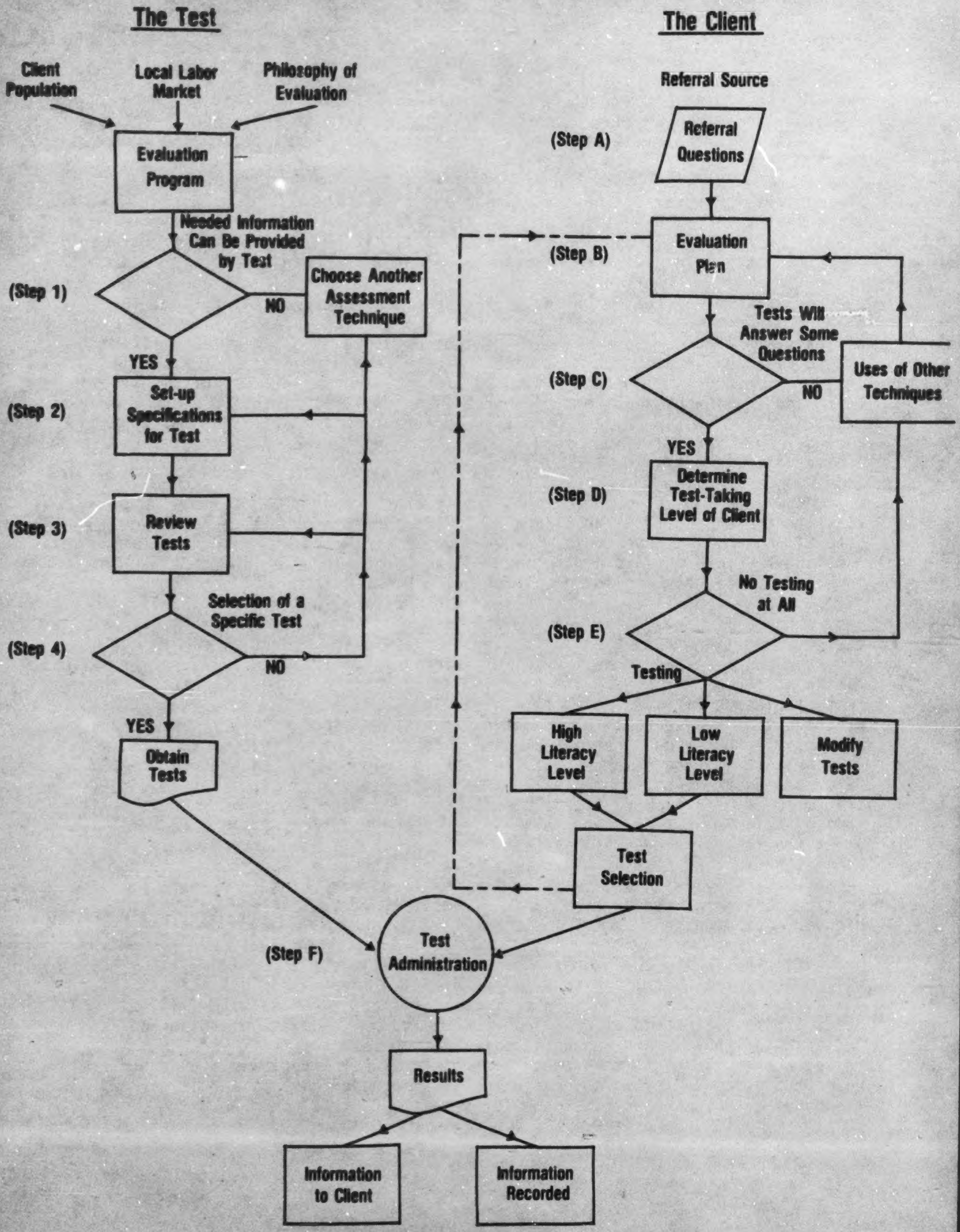


Figure 1: A Model for Test Use

In testing a specific client (right side of Figure 1), the process begins with the questions asked by the referral source (Step A). If these questions are precise, then a detailed evaluation plan can be developed upon entering the facility. If not, the evaluator will have to obtain additional information prior to developing the plan. The evaluation plan (Step B) is established to answer specific questions about the client and during this phase, questions of what techniques to use are not asked. Once specific needs are established, the first decisions revolve around what techniques will best answer the referral questions. With regard to testing, the first question is, "Will any testing be necessary to assess the client" (Step C)? If the answer is "no," then other techniques are used.

If the answer is "yes," then the next phase is to determine the test taking ability of the client (Step D). Some suggestions for this are given on page 6 and need not be repeated. Once the client's test taking skills, mainly literacy in the English language, have been determined, some decisions can be made (Step E). If the client does not have minimal test taking skills, other techniques must be used and the evaluation plan revised to reflect this. While literacy or test taking ability obviously occur on a continuum, Figure 1 has simplified this continuum by giving only three testing options. The first is for "high literacy" clients, a term used to indicate a client who can take almost any standardized test. The second, "low literacy," implies the use of nonreading, nonverbal, or low reading tests. The third choice is to modify specific tests to meet client needs [see Botterbusch (1976) and Dickson (1976)]. The evaluator selects the actual test based on the information needed and the client's test taking skill level. This test selection is recorded on the evaluation plan (the dotted line on Figure 1).

Actual test administration is the next step (F). Here the client and the selected test are merged. If possible, the test should be administered according to the instructions in its manual, and any modifications must be carefully made. Even though the test and client have been carefully chosen for each other, there still can be unforeseen problems. The evaluator must be aware of these and make sure the client knows what the instructions mean, that the client is not too anxious to take the test, etc. Following administration and scoring, the results are recorded and shared with the client. The client must understand to the best of his ability what the results mean, how they will reflect upon the results of the rest of the evaluation, and especially, how they may affect the client's future goals.

Part I has attempted to provide the evaluator with some general guidelines on how to properly use psychological tests within the evaluation program. It is realized that some of these suggestions, recommendations and procedures will appear general to the reader. There is a reason for this--each facility is a unique combination of clients, programs, referral sources, staff and financial resources. In addition, each facility exists in a unique community. To set forth highly regimented guidelines for test use would be impractical to the evaluator and place the writer in a position of assuming that he knows every facility and its procedures. The ultimate responsibility for the appropriate use of tests depends upon the professional judgment of the evaluator. The first part of this publication provided some general information on testing; the second part provides specific information on tests and gives reference sources. It is hoped that these two parts will help and encourage the evaluator to make appropriate decisions about test use and selection.

PART II

The second part of this publication contains reviews of tests that are either widely used in vocational evaluation or tests that have potential for widespread use in client assessment. The description of each instrument is presented in a standardized format, designed to objectively summarize each device. The format and what is contained in each category follows this introduction. In using this format, the evaluator should keep in mind that only very basic information is presented and that he should consult additional sources prior to selecting a test for his use.

Test Review Outline

The name of the instrument, as listed in the Mental Measurements Yearbook and on the test manual, is centered in capital letters on the top of the page. Any common names or abbreviations are given in parentheses following the formal name. The next line contains the form(s) of the test, if any, which were reviewed for this publication. The review applies only to the form(s) listed on this second line. The most recent copyright date is placed in parentheses on the right margin.

1. Purpose - Why is this instrument administered? The specific traits, aptitudes, abilities, skills, etc. that the test is designed to measure are presented here.

2. Final Score - The final scores are identified and the method of presentation given.

3. Description

    A. Administration - Individual or group test, pencil-and-paper or apparatus, administration time, and any required special conditions or materials are presented.

    B. Content - Content includes descriptions of the number of items, item format, and the type of item.

    C. Scoring - This contains the time required for scoring, the availability of computerized scoring services, and the use of conversion tables.

4. Materials Required - A list of the materials necessary to administer and score the test is given.

5. Appropriate Groups - The necessary physical and "mental" skills necessary to successfully complete the test are indicated.

6. Technical Considerations

    A. Norm Groups - On what groups are norms available? How are the groups described in the test manual?

    B. Reliability - This presents typical reliability data found in the test manual.

C. Validity - What basis does the test have for what it claims to measure?

7. APA Level - See page 94.

8. Sources of Information - All Buro's reviews are given. For example, "MMY 7:351" refers to the Seventh Mental Measurement Yearbook, review number 351. The books listed on pages were searched and descriptions and discussions of the tests contained in these books are included. "Cronbach, 528," for example, tells that the book by L. J. Cronbach contains information about the test on page 528. Some of the more recently published tests have not been reviewed in the MMY's or by other publishers. It must also be noted that some tests have been extensively revised and/or new forms have been added. In cases where this occurs, the references for earlier forms have been given; these have been indicated.

9. Available From - The name of the test publisher is given. See page 95. for addresses.

10. Comments - General critical comments are presented here.

ADULT BASIC LEARNING EXAMINATION (ABLE)

Forms A and B (1971)

1. Purpose - The ABLE is designed "to measure the level of educational achievement among adults" who have not completed a formal education. In addition to being a method for determining literacy level, it can be used "in evaluating a number of efforts to raise the educational level of these adults."

2. Final Score - The ABLE yields six grade-equivalent scores: three in verbal skills (vocabulary, reading, and spelling) and three in arithmetic (computation, problem solving, and total).

3. Description

   A. Administration - There are three levels to the ABLE: Level I (grades 1-4), Level II (grades 5-8), and Level III (grades 9-12). The examiner selects the appropriate level for the examinee. Total testing time for Levels I and II is about two and one-half hours; Level III testing time is over three hours. Levels I and II are not timed. Although Level III tests are timed, the manual states that the time limits are for administrative convenience only. Some tests are dictated: Level I - vocabulary, spelling, and problem solving; Level II - vocabulary and spelling; Level III - vocabulary. This is a group administered pencil and paper test and requires no special arrangements. At Levels I and II there is a choice of non-reuseable test booklets or machine answer sheets; Level III provides for two separate answer sheets which can be hand or machine scored.

   B. Content - There are five sub-tests at each level: vocabulary, reading, spelling, arithmetic computation, and problem solving. Although item formats differ with sub-tests and level, each level contains multiple-choice items and items which require the writing of a word or the answer to an arithmetic problem.

   C. Scoring - Levels I and II, spelling and arithmetic computation must be hand scored; the remaining Level I and II tests and all Level III tests may be hand or machine scored. Templates are used in hand scoring and the number of correct responses in each section are counted and recorded. Raw scores are compared to norm tables. At Levels I and II the scores are placed on a graph on the last page of the test booklet. Level III has a separate test score recording sheet. Hand scoring takes about ten minutes.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Manual

   D. Scoring templates

   E. Separate answer sheet for Level III; optional answer sheets for Levels I and II.

5. Appropriate Groups - The wide range of achievement levels covered by ABLE and the adult item content imply that the ABLE could be used as a measure of literacy for almost all clients whose vision and writing skills permit successful completion of the test. The tests, however, are fairly lengthy and it is urged that the sub-tests be interspersed with break periods.

6. Technical Considerations

   A. Norm Groups - The ABLE was developed and normed on 1,000 students (grades 2 through 7), Job Corps trainees, and adults enrolled in basic education programs. The grade equivalent norms were developed on high school students. All samples are carefully defined.

   B. Reliability - Split-half reliabilities for the sub-tests at all levels range from .60 to .96. Most of the reported reliability coefficients are unusually high.

   C. Validity - Although the manual gives no predictive validity studies, it does provide correlations with the Stanford Achievement Test. Most of these are in the .50's and .60's.

7. APA Level - A

8. Sources of Information - MMY 8:2; MMY 7:3; Anastasi, 408-410; Bolton, 93.

9. Available From - The Psychological Corporation

10. Comments - The ABLE can be used to determine the achievement of adults lacking in formal education. Unlike many reading and arithmetic achievement tests, the ABLE was specifically designed for adults having little formal education.

CALIFORNIA ACHIEVEMENT TESTS (CAT)

Forms C and D; Levels 10-19 (1977)

1. Purpose - This series of achievement tests is designed to measure academic skills in four major content areas: reading, spelling, language, and mathematics. The CAT series is intended to be used for the large scale achievement testing of school populations. Except for Level 10, this test is totally new.

2. Final Score - The CAT includes numerous final scores which differ with the level of the test. The upper levels (14-19) of the test contain the following scores: reading vocabulary, reading comprehension, total reading, spelling, language mechanics, language expression, total language, mathematics computation, mathematics concepts and applications, total mathematics, total battery, and reference skills. These scores are presented in the grade equivalents, scale scores, percentile ranks, and stanines. In addition, mastery scores are presented in the content areas measured by the specific test. Some examples of these are: recall of facts, addition, number set theory, and verbs.

3. Description

   A. Administration - The CAT series is a group administered pencil-and paper-test with separate test booklets at each level. The level, its grade equivalent, and the total estimated administration times (tests that are not read by the examiner are timed) are given below:

Level	Grade Equivalent	Total Administration Time
10	K.0-1.3	2 hours 41 minutes
11	K.5-1.9	2 hours 12 minutes
12	1.5-2.9	3 hours 5 minutes
13	2.5-3.9	3 hours 16 minutes
14	3.5-4.9	3 hours 20 minutes
15	4.5-5.9	
16	5.5-6.9	
17	6.5-7.9	
18	7.5-9.9	
19	9.5-12.9	

Because they have the same number of sub-tests and format, Levels 14-19 can be administered during the same test session. In order to determine what level to administer, the examiner first gives one of the two locator tests (grades 1-6; 6-12) which contain a total of 40 vocabulary and mathematical items. The level is selected from the results of this preliminary measure. In addition to the locator tests, there are separate practice tests that should be given a day or two before the battery. All responses are made on a separate answer sheet.

B. Content - All test items are multiple-choice with between three and five alternatives. At the lower levels, the examiner reads many of the vocabulary and reading sections. Levels 14-19 contain the following seven tests and item numbers: reading vocabulary-30 items; reading comprehension-40 items; spelling-20 items; language mechanics-25 items; language expression-38 items; mathematics computation-40 items; mathematics concepts and applications-45 items; and reference skills-25 items.

C. Scoring - Two methods of hand scoring are available for Levels 14-19. The first uses a separate IBM answer sheet and stencil; the second is a form concealed key on the inside of a folded answer sheet. Raw scores are compared to norms tables. Hand scoring is estimated to take about ten minutes per test. Computer scoring is available.

4. Materials Required

   A. Locator test

   B. Practice test

   C. Test booklet

   D. Two pencils

   E. Answer sheet

   F. Scoring stencil

   G. Examiner's manual

   H. Norms tables

5. Appropriate Groups - Because the CAT is aimed at in-school age groups, much of the item content at the lower level is for younger children. The content of the upper levels is appropriate for high school age persons. This series could be used as a literacy test for almost all clients whose vision and writing skills permit taking it.

6. Technical Considerations

   A. Norm Groups - National norms were developed on over 177,000 students in grades kindergarten through grade 12. These students were tested twice--in the fall of 1976 and the spring of 1977. The sampling techniques are excellent and are described in the test coordinator's handbook. In addition to norms, the CAT is also based on criterion-referenced material.

   B. Reliability - The CAT has three types of reliability data: (1) internal consistency (KR-20) results, mostly in the .80's; (2) test-retest over a two to three week period, mostly in the .70's; and (3) alternate forms, six months, mostly in the .60's. All of these results are very acceptable.

C. Validity - There are two major sources of validity. First, the content of the CAT was based on a careful analysis of school subject matter. Second, regressions and cross-validation designs with other tests yielded promising results. The CAT now needs to be validated against criteria such as school grades.

7. APA Level - A

8. Sources of Information - (earlier editions) MMY 8:10; MMY 7:5; MMY 6:3; MMY 5:2; MMY 4:2; Noll & Scannell, 164; Barnette, 54-58.

9. Available From - CTB/McGraw-Hill

10. Comments - The CAT is a well developed, well designed test intended for use in education both as a student diagnostic test and as an institution program evaluation device. Its use with a rehabilitation population would be to measure the basic literacy of an individual. The need for the locator tests and the recommended use of a practice test means that the user will have to carefully schedule the client's time. Some of the item content at the lower levels is too childish for adults. This test could be best used when a thorough evaluation of the literacy skills of a young person are needed.

GRAY ORAL READING TEST

Forms A, B, C, and D (1963)

1. Purpose - The two major functions of this test are: "First, to provide an objective measure of growth in oral reading from early first grade to college, and second, to aid in the diagnosis of oral reading difficulties."

2. Final Score - A grade 1 through 12 equivalent score for oral reading is the only reported score.

3. Description

   A. Administration - The testing material is individually administered to the examinee who reads short passages aloud. The examinee reads passages until he makes seven or more errors of two successive passages. Following the completion of each passage the examinee is asked four questions about the content of each. There are eight different types of reading errors for each passage. The test takes about 20 minutes to administer.

   B. Content - The 13 passages range in difficulty from grade 1 through adult. Each passage is more difficult than the preceding one.

   C. Scoring - Each passage is scored separately. The number of seconds required to read each passage and the number of errors made in each passage are combined according to a chart to produce a final passage score. Passage scores are added to produce a total Passage Score.

4. Materials Required

   A. Booklet with reading passages

   B. Manual

   C. Record booklet

   D. Stopwatch

5. Appropriate Groups - Because the test contains a wide range of material and is individually administered, it is applicable for the minimally literate as well as highly literate persons.

6. Technical Considerations

   A. Norm Groups - The test has "tentative norms" based on about 500 male and female students--about 40 in each grade 1 through 12. Separate norms for each sex are available. The selection of the norm group was not based on a systematic sampling design and is too small to provide an adequate norm base.

   B. Reliability - Reliability information is given mainly in terms of standard errors of measurement. These errors are very small as compared with other reading tests. Equivalent form reliability is in the high .90's.

C. Validity - The developer of the test bases his validity on the content, i.e., the material used in the passages, and therefore does not present any empirical evidence of validity.

7. APA Level - B

8. Sources of Information - MMY 6:842.

9. Available From - The Bobbs-Merrill Company

10. Comments - The Gray is designed to assess for oral reading level. Although not scored, the test presents clear definitions of reading errors and encourages the use of behavioral observation and interpretation. The major problem with the Gray is its lack of adequate norm groups. In spite of this major problem, the Gray is useful for the partial assessment of oral reading skills.

NELSON-DENNY READING TEST

Forms C and D (1976)

1. Purpose - This test is designed to assess the reading ability of high school and college level students and adults, and can be used for screening, predicting academic success, and for a broad diagnosis of individual reading problems.

2. Final Score - The test yields four final scores: (1) vocabulary, (2) comprehension, (3) total, and (4) reading rate. Scores can be given in percentiles, grade equivalent scores, and standard scores.

3. Description

   A. Administration - This pencil-and-paper test is group administered. Under normal test administration conditions, the test can be given in about 35 minutes. When testing adults or college students, the time limits under some conditions can be reduced; this reduces the total test administration time to between 25 and 30 minutes. A separate answer sheet is used. The test is given under somewhat speeded conditions.

   B. Content - The test consists of two parts. The first is a 100 item vocabulary test; the second is eight passages with 36 items about these passages. All items are five alternative multiple-choice.

   C. Scoring - There are four different answer sheets available: (1) MRC answer cards for electronic scoring, (2) self-marking answer sheets with a carbon center that eliminates the need for a separate scoring key, (3) IBM answer sheets, and (4) Digitek answer sheets. The last two can be used for either machine or hand scoring. The manual does not contain any estimates of scoring time. However, the self-scoring sheets should take only a few minutes to score and can be scored by the client. Hand scoring with stencils on the IBM or Digitek will take longer. Scores are entered on a profile sheet.

4. Materials Required

   A. Test booklet

   B. Answer sheet or card

   C. Two pencils

   D. Scoring stencil

   E. Profile sheet

   F. Stopwatch

   G. Clock

5. Appropriate Groups - Because of the difficulty of the test items and the use of a separate answer sheet, it is suggested that the device only be used for clients who are test sophisticated and who function at a fairly high level. The items are academically oriented and, therefore, are biased in favor of people with middle-class backgrounds.

6. Technical Considerations

   A. Norm Groups - The test has excellent norms based on a group of over 15,000 high school students. This national sample is well defined. College norms are also available; however, the publisher clearly warns the reader that these are not based on a random sample. (It is the reviewer's opinion that the college norms are more complete than the norms of many tests that claim to have been based on national random samples.) Adult norms for the shortened administration time are also available. Separate norms are given for high school grades 9 through 12 and college freshman through senior.

   B. Reliability - Alternate form reliabilities range from .54 to .91. Split-half reliability estimates using the Spearman-Brown formula range from .81 to .98. The alternate form reliabilities are acceptable; the high split-half estimates reflect the formula used and the speeded condition of the test. Standard error of measurement data are also provided.

   C. Validity - Some validity data are contained in the manual--correlations with specific course grades and overall GPA, intercorrelations of tests scales on test-retest study, and correlations with two other achievement tests are available. These results provide some good evidence for the overall validity of the test.

7. APA Level - A

8. Sources of Information - MMY 8:735; (earlier editions) MMY 6:800; MMY 4:544; Horrocks and Schoonover, 164-165.

9. Available From - Houghton Mifflin

10. Comments - The instrument is well designed, thoroughly researched, and adequately normed. It can be used for investigating the reading level of clients who are considering future education or training which requires traditional academic skills.

PEABODY INDIVIDUAL ACHIEVEMENT TEST (PIAT)

(1970)

1. Purpose - The PIAT is intended to "provide a wide range, screening measure of school achievement in the areas of mathematics, reading, spelling, and general information."

2. Final Score - The test yields six final scores: (1) mathematics, (2) reading recognition, (3) reading comprehension, (4) spelling, (5) general information, and (6) total.

3. Description

   A. Administration - The PIAT is individually administered in from 30 to 40 minutes. None of the five sub-tests (corresponding to scores one through five above) are timed; items are presented orally and the examinee responds by selecting the appropriate word, number, or illustration from the four alternatives.

   B. Content - Items for all tests are multiple-choice. Eighty-four items are contained in each sub-test: Mathematics; Reading Recognition; Reading Comprehension; Spelling; and General Information. The items are contained in two flip-books.

   C. Scoring - Sub-tests are scored according to the number of correct items between a basal (five consecutive correct responses) and a ceiling (five errors in seven consecutive responses). Scores are converted into a variety of norms (grade scores, percentile ranks, age scores, and standard scores). Results are plotted on a profile sheet.

4. Materials Required

   A. Test booklets

   B. Manual

   C. Individual record booklet

5. Appropriate Groups - Because the PIAT is administered orally, it is appropriate for people who cannot write. The wide range of items and the versatility of test administration makes the test useful for individuals having a kindergarten through twelfth grade achievement level. Since the test is individually administered, it can be used for people who do not perform well on group tests for emotional, cultural, or intellectual reasons.

6. Technical Considerations

   A. Norm Groups - The PIAT was normed on 200 boys and girls in each grade (kindergarten through 12), a total of about 3,000 students in all. These students were selected from different geographical regions. The manual clearly describes the sources and characteristics of the samples.

B. Reliability - Test-retest reliabilities for the sub-tests range between .52 and .94 - most are well within the acceptable range. The total test scores have coefficients ranging from .82 to .92.

C. Validity - Content and concurrent validity data are presented. Careful item analysis and selection procedures were used to assure the content validity of the items. Concurrent validity is given in terms of correlations with other achievement and intelligence tests.

7. APA Level - B

8. Sources of Information - MMY 8:24; MMY 7:17; Bolton, 93.

9. Available From - American Guidance Service

10. Comments - The PIAT could be used in evaluation facilities for testing clients who are resistant to group tests and as a possible method for establishing rapport with a client who is suspicious or hostile of testing. The PIAT could also be used as an opportunity to make behavioral observations of a clinical nature. It is this reviewer's opinion that the PIAT deserves the widespread use it has received.

TESTS OF ADULT BASIC EDUCATION (TABE)

Forms 3 and 4; Levels E, M, D (1976)

1. Purpose - This series of tests is designed to measure an adult's literacy skills in the areas of reading, mathematics, and language. According to the manuals, the TABE can be used for several purposes: to assess achievement in basic skills; to identify areas of weakness in these skills; and to measure growth in these skills after instruction.

2. Final Score - The TABE yields several final scores. Level E contains six scores: total reading (combined from vocabulary and comprehension scores) and total mathematics (combined from computation and concepts and problems scores). Levels M and D give eight scores: total reading (combined from vocabulary and comprehension scores); total mathematics (combined from computation and concepts and problems); mechanics and expression and spelling. (These last two scores are the language part of the test.) In addition, all levels give a total battery score. Scores are reported as scale scores and grade equivalent scores.

3. Description

   A. Administration - The TABE series is a group administered pencil-and-paper test with separate test booklets for each level. The three levels, the approximate grade equivalent, and total estimated administration times are as follows:

Level	Grade Equivalent	Total Administration Time
E (Easy)	2.5-4.9	2 hours 7 minutes
M (Medium)	4.5-6.9	3 hours 29 minutes
D (Difficult)	6.5-8.9	3 hours 11 minutes

   All tests are timed except a Level E vocabulary test which is read by the examiner. In order to determine what test level the client should take, a combination practice exercise and locator test (20 vocabulary and 18 mathematics multiple-choice items) is gven first. Answers are made on a separate answer sheet or in the test booklet.

   B. Content - Most test items are multiple-choice with four or five alternatives. Other items require a true or false answer. Each of the three levels is composed of sub-tests, for example, Level M contains the following ten tests and item numbers: vocabulary-40 items; comprehension-42 items; computation-48 items; fractions-20 items; concepts-25 items; problems-15 items; capitalization-38 items; punctuation-42 items; expression-29 items; and spelling-32 items.

   C. Scoring - The test may be scored in three ways. If answers have been marked in the test booklet, an answer key is used. Answer sheets are scored using overlay scoring stencils. Special answer sheets on which "the marked responses show up on a concealed key on the inside of the answer sheet" are the third way of scoring. Conversion tables are used to give grade equivalent and scale scores. It is estimated that scoring takes under ten minutes per examinee. No machine scoring is available.

4. Materials Required

   A. Practice exercise and locator test

   B. Test booklet

   C. Answer sheets

   D. Two pencils

   E. Examiner's manual

   F. Scoring stencil

5. Appropriate Groups - The TABE is intended for use by adults in adult education programs. Special care was given to make all the items free of cultural bias and to make the content appropriate for adults. Because the TABE is a literacy test, reading level is not appropriate. Because oral communication is required for Level E, the test would have to be modified for deaf persons. Otherwise, visual acuity and the ability to use the answer sheet are required.

6. Technical Considerations

   A. Norm Groups - The TABE norms are based equating the TABE with the 1970 edition of the California Achievement Tests. Both tests were administered to 18,183 students in grades 2 through 9 in 36 states.

   B. Reliability - Internal consistency (KR-20) reliabilities for each level and form are in the .80's and .90's. Test-retest reliabilities range from .60 to .85. These estimates are very adequate.

   C. Validity - The only real validity information compares TABE results with passing or failing the GED test in which the TABE proved to accurately predict GED results. This single study, while encouraging, does not establish the TABE as a thoroughly validated test.

7. APA Level - A

8. Sources of Information - MMY 8:33; (earlier editions) MMY 7:32.

9. Available From - CTB/McGraw-Hill

10. Comments - The TABE appears to be a well designed, well developed test for assessing the basic literacy skills required by an adult. The use of the locator test provides an excellent method for determining the appropriate level.

EDWARDS PERSONALITY INVENTORY (EPI)

Introductory Edition (1967)

1. Purpose - The EPI "was designed to measure a large number of personality characteristics in which normal individuals vary." The manual sees the EPI as being useful in counseling, prediction, and research.

2. Final Score - The EPI provides measures of 53 specific personality variables. These final scores are presented in percentiles. The scales and the booklets containing each scale is as follows:

   a) Booklets 1A and 1B. 14 scores: plans and organizes things, intellectually oriented, persistent, self-confident, has cultural interests, enjoys being the center of attention, carefree, conforms, is a leader, kind to others, worries about making a good impression on others, seeks new experiences, likes to be alone, interested in the behavior of others.

   b) Booklet 2. 11 scores: anxious about his performance, avoids facing problems, is a perfectionist, absentminded, sensitive to criticism, likes a set routine, wants sympathy, avoids arguments, conceals his feelings, easily influenced, feels misunderstood.

   c) Booklet 3. 15 scores: motivated to succeed, impressed by status, desires recognition, plans work efficiently, cooperative, competitive, articulate, feels superior, logical, assumes responsibility, self-centered, makes friends easily, independent in his opinions, is a hard worker, neat in dress.

   d) Booklet 4. 13 scores: self-critical, critical of others, active, talks about himself, becomes angry, helps others, careful about his possessions, understands himself, considerate, dependent, shy, informed about current affairs, virtuous.

3. Description

   A. Administration - This pencil-and-paper test is untimed and can be administered to either individuals or groups. A separate answer sheet is used. The manual estimates that each booklet can be completed in about 40 minutes. However, some rehabilitation clients may take longer. Because the instructions for all booklets are the same, it is "possible to administer different booklets simultaneously in a group setting." (Because booklets 1A and 1B are alternate forms, only one of the two is administered; in actual usage, four separate booklets are given.)

   B. Content - Each of the five booklets contain 300 items; the number of items on each scale ranges from 10 to 58. All items are written in the same style: (1) begin with "He," (2) a verb, and (3) describes a behavior or characteristic of that person (e.g., "He plans his work carefully"). The examinee responds with a true or false answer. The instructions ask the examinee to describe himself as he believes persons closest to him see him.

C. Scoring - Tests may be either hand or machine scored. There are two hand scoring methods: (1) a self-scoring answer sheet with a carbon insert and (2) an IBM answer sheet with scoring stencils. A separate answer sheet is needed for each booklet. Scoring time for all four booklets is estimated to be 20 minutes. In hand scoring, norm tables are used to convert raw scores into percentiles.

4. Materials Required

   A. Four test booklets

   B. Four answer sheets

   C. Pencils

   D. Administration manual

   E. Scoring stencils

5. Appropriate Groups - While the EPI manual contains no estimate of the reading level required, this reviewer estimates that a sixth grade level of comprehension is required. The examinee must also be able to use a separate answer sheet. The EPI is designed for "normal"; therefore, it should not be used to diagnose persons with severe psychological or emotional problems. The large number of scales make the EPI useful for pinpointing a particular behavior problem.

6. Technical Considerations

   A. Norm Groups - The EPI was normed on 211 male and 338 female students at the University of Washington and on 111 male and 163 female Seattle high school juniors. No national norms or special groups exist. The following percentile norms are available in the manual: male and female college students and male and female high school students.

   B. Reliability - Kuder-Richardson 20 (internal consistency) values for the 53 scales were calculated; these range from .69 to .95 with most being in the .80's. Alternate form reliabilities using Booklets 1A and 1B between .60 and .86 were found. While these are adequate, the manual contains no data on test-retest reliability. However, the scales must be considered very homogeneous.

   C. Validity - The manual contains no validity data.

7. APA Level - C

8. Sources of Information - MMY 7:73; Cronbach, 557.

9. Available From - Science Research Associates

10. Comments - The large number of scales that are at least to some degree behaviorally orientated make the EPI appealing to the evaluator who wishes to pinpoint behavior/personality problems in a relatively "normal" client. In this context the EPI could be used as a measure of success in counseling or in work adjustment programs. Some women may have problems relating to

the use of the word "he" in every item. The problems in using the EPI are mostly technical--the test has a totally inadequate norm base and contains not even a hint of validity data. Any user of this test must carefully weigh these severe limitations against its potential use.

EDWARDS PERSONAL PREFERENCE SCHEDULE (EPPS)

(1959)

1. Purpose - The EPPS "was designed as an instrument for research and counseling purposes, to provide quick and convenient measures of a number of relatively independent normal personality variables." The selection of these variables is based on the list of manifest needs developed by H. A. Murrey and others.

2. Final Score - The EPPS provides measures of the following personality variables: (1) achievement, (2) deference, (3) order, (4) exhibition, (5) autonomy, (6) affiliation, (7) intraception, (8) succorance, (9) dominance, (10) abasement, (11) nurturance, (12) change, (13) endurance, (14) heterosexuality, and (15) aggression. Final scores are given in percentiles.

3. Description

   A. Administration - This pencil-and-paper test can be individually or group administered. The EPPS is not timed. The manual states that most college students can finish the test in 50 minutes. The time may be longer for rehabilitation clients. A separate answer sheet is used.

   B. Content - The test contains 225 items; each forced-choice item contains two statements and the examinee selects the statement that best describes him. All of the statements begin with "I"; most use the verb "like" (e.g., "I like to help my friends when they are in trouble."). Items have been carefully constructed to minimize the influence of social desirability.

   C. Scoring - Scoring is either by hand or machine. Raw scores for each of these 15 scales are converted to percentiles, and these are presented in profile fashion. Special self-scoring answer sheets are used for hand scoring. An IBM answer sheet and scoring stencils can also be used. Scoring time should be less than ten minutes.

4. Materials Required

   A. Test booklet

   B. Answer sheet

   C. Pencils

   D. Manual

   E. Scoring stencils

5. Appropriate Groups - The EPPS was developed for college students and adults; therefore, it has a fairly high reading level. The instrument is designed for normal persons and is not appropriate for use with severely disturbed clients. A fair amount of test sophistication is necessary to use the separate answer sheet.

6. Technical Considerations

   A. Norm Groups - The EPPS was normed on the following groups: (1) 760 male college students, (2) 749 female college students, (3) 4,031 adult males, and (4) 4,932 adult females. The sample characteristics and sampling plan for these groups are not clearly defined.

   B. Reliability - The manual reports split-half and test-retest reliabilities for each scale. Split-half reliabilities range from .60 to .87; test-retest from .74 to .86. Considering the small number of items in each scale, these reliabilities are quite acceptable.

   C. Validity - The manual reports little validity data, except for a few correlations with other personality tests. It should be mentioned that the EPPS has been used in a wide variety of research studies and that the author of the EPPS has not revised his manual to include this new material.

7. APA Level - C

8. Sources of Information - MMY 8:542; MMY 7:72; MMY 6:87; MMY 5:47; Anastasi, 510-513, 518, 519; Cronbach, 551, 557; Guion, 329-330; Horrocks and Schoonover, 405-411; Maloney and Ward, 328; Thorndike and Hagen, 687; Super and Crites, 537-555.

9. Available From - The Psychological Corporation

10. Comments - The EPPS was designed for counseling essentially normal students and the instrument is adequate if used for this purpose. The high reading level required for its successful completion may hinder its use for many clients.

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI)

Individual Form, Old Group Form, New Group Form (1967)

1. Purpose - The MMPI "is designed to provide an objective assessment of some of the the major personality characteristics that affect personal and social adjustment. The point of view determining the importance of a trait in this case is that of the clinical . . . worker who wishes to assay those traits that are commonly characteristic of disabling psychological abnormality."

2. Final Score - The instrument contains ten clinical scales: (1) hypochondriasis, (2) depression, (3) hysteria, (4) psychopathic deviate, (5) masculinity-feminity, (6) paranoia, (7) psychasthenic, (8) schizophrenia, (9) hypomania, and (10) social. In addition, there are three validation scales. Results are presented in scale scores and are also plotted.

3. Description

    A. Administration - The device is available in three separate formats: (1) the Individual Form consists of a box containing 550 statements printed separately on small cards. The examinee sits at a table and sorts each statement into true or false piles. (2) The Old Group Form ("The Booklet Form") is a 566 item paper-and-pencil test using a separate answer sheet and a booklet containing the items. (3) The New Group Form ("Form R") uses the first 399 items of the Old Group Form to produce the basic scores; this form presents the items in a lap-book with step-down pages. A separate answer sheet is used. For both group forms, the client responds by marking the items "true" or "false." The test is untimed; administration time varies between 45 to 90 minutes. The MMPI can be administered individually or in groups.

    B. Content - Each item contains a brief statement about a personal characteristic, personality trait, feeling, habit, etc. The examinee marks each answer as being true or false.

    C. Scoring - Scoring may be done by hand or machine. Because of the number of scales and the availability of computer interpretation, machine scoring is recommended. For hand scoring one key is used for each scale; raw scores are converted with the appropriate norm tables to T scores. Profile plotting and coding are also used. Hand scoring time is estimated to take about 45 minutes.

4. Materials Required (for booklet forms)

    A. Booklet

    B. Answer sheet

    C. Two pencils

    D. Manual

E. Handbook

F. Atlas

G. Scoring keys

H. Record sheets

5. Appropriate Groups - According to the manual, clients 16 years or older with a sixth grade reading level can be expected to complete the MMPI without difficulty. Because of the nature of some of the items, the examiner should determine beforehand whether the instrument is appropriate. The MMPI should not be routinely given to all clients.

6. Technical Considerations

   A. Norm Groups - Originally, the MMPI was developed by contrasting scores of about 700 "normal" adults with 800 carefully studied clinical cases.

   B. Reliability - The manual presents the results of three test-retest reliability investigations. Scale scores ranged from .46 to .91, with most of the coefficients in the .70's.

   C. Validity - The predictive validity of the MMPI was originally determined by comparing it with clinical diagnosis. More than 60% of the cases were successfully predicted. Since that time entire volumes have been written on the MMPI. The research evidence available on the MMPI is too massive to be summarized here.

7. APA Level - C

8. Sources of Information - MMY 8:616; MMY 8:617; MMY 8:618; MMY 8:619; MMY 8:620; MMY 8:621; MMY 8:622; MMY 8:623; MMY 8:624; MMY 7:104; MMY 7:105; MMY 7:106; MMY 7:107; MMY 7:108; MMY 7:109; MMY 6:143; MMY 5:86; MMY 4:71; MMY 3:60; Anastasi, 497-504, 521; Barnette, 83-85, 95-99, 167-174, 209-211; Bolton, 111-112; Cronbach, 527-548; Guion, 315-316; Jackson & Messick, 524-526, 548, 550; Maloney & Ward, 26-27, 311-313, 331-343; Super & Crites, 520-537; Thorndike & Hagen, 404-407.

9. Available From - The Psychological Corporation

10. Comments - The MMPI has a long history of being a valuable and trusted tool in the diagnosis of "abnormal" behavior. Graduate training in psychology and in the use of the MMPI is necessary before the examiner can use the test. The MMPI is often used when the professional suspects that a client may have a serious psychological problem.

SIXTEEN PERSONALITY FACTOR QUESTIONNAIRE (16PF)

Form E (1976)

1. Purpose - The 16PF is intended to provide "the most complete coverage of personality possible in a brief testing time." Form E is a new low literate form "for use with persons of limited educational and cultural background." The factors measured by Form E are the same as those measured by other 16PF forms.

2. Final Score - Scores are given on 16 bipolar primary factors: (1) reserved - outgoing, (2) less - more intelligent, (3) affected by feelings - emotionally stable, (4) humble - assertive, (5) sober - happy-go-lucky, (6) expedient - conscientious, (7) shy - venturesome, (8) tough - tender-minded, (9) trusting - suspicious, (10) practical - imaginative, (11) forthright - shrewd, (1) self-assured - apprehensive, (13) conservative - experimentary, (14) group dependent - self-sufficient, (15) undisciplined self - conflict-controlled, and (16) relaxed - tense. In addition, there are six second-order factors: (1) adjustment - anxiety, (2) introversion - extroversion, (3) tender-minded emotionally - alert poise, (4) subduedness - independence, (5) naturalness - discreetness, and (6) cool realism - prodigal subjectivity. The profile sheet provides space for scoring and plotting the 16 primary factors. Scores are reported in standard ten scores (sten).

3. Description

   A. Administration - Form E can be administered either to individuals or to groups. The test is untimed and the manual contains no estimate of the average completion time. Answers are recorded on a separate sheet. The developer estimates that a third or fourth grade reading level is required; a tape recorded version can be used for those reading below the third grade level.

   B. Content - The 16PF contains 128 forced-choice items such as "would you rather play baseball or go fishing." The person selects the activity, feeling, preference, etc. that he would rather do or be.

   C. Scoring - The test can be hand scored with a clear plastic key. Raw scores are found for each of the 16 scales; these are converted to stens using the appropriate norms tables. Results are plotted either on the back of the answer sheet or on a separate profile sheet. Computer scoring is also available.

4. Materials Required

   A. Test booklet

   B. Answer sheet

   C. Pencils

   D. Test manual

    E. Scoring stencil

    F. Handbook for interpretation

    G. Profile sheet (optional)

5. Appropriate Groups - The 16PF can be readily used to assess persons with low reading skills in English. The items are printed in large type and placed in an easy to read format. The taped version of the test could extend its use to persons who are illiterate and those with visual handicaps. A highly educated person may be able to "see through" the item content, thus making the 16PF easy to fake.

6. Technical Considerations

    A. Norm Groups - Separate norms are on males and females ages 14-20 and 21-29, prison inmates, schizophrenics, males and females with severe hearing losses, and culturally disadvantaged males. Because many of the sample characteristics, including sample size, are not given in the manual, the adequacy of the samples cannot be judged.

    B. Reliability - Internal consistency estimates are given for each of the 16 scales--correlations range from -.12 to .71. The reliability of this form of the 16PF has not been firmly established. The data reported in the manual are somewhat disappointing.

    C. Validity - Validity is based on construct validity with pure factor scores and with a factor analysis of items. These results provide some good evidence that the 16PF contains scales that are independent of each other.

7. APA Level - C

8. Sources of Information - MMY 8:679; (earlier editions) MMY 7:139; MMY 6:174; MMY 5:112; MMY 4:87; Anastasi, 509-510; Barnette, 180-185; Bolton, 113-114; Cronbach, 559; Guion, 320; Thorndike & Hagen, 690.

9. Available From - Institute for Personality and Ability Testing

10. Comments - This form of the 16PF will provide the user with a tool to assess many common personality traits in persons with low literacy skills. Considering high reading level of many personality tests, the literacy level alone is a positive point for the 16PF. Because the technical aspects of the test need additional clarification, the form should be used with caution.

WORK ENVIRONMENT PREFERENCE SCHEDULE (WEPS)

(1973)

1. Purpose - "The WEPS is designed to measure a personality construct, 'bureaucratic orientation,'* which reflects a commitment to the set of attitudes, values, and behaviors that are characteristically fostered and rewarded by bureaucratic organizations." The title of the test is somewhat misleading--the WEPS measures only the personality construct of accepting work in a bureaucracy.

2. Final Score - The WEPS gives one final score in percentile form.

3. Description

   A. Administration - The WEPS is a self-administering pencil and paper test in which the person marks the answers on the test sheet. Although not specifically stated in the manual, the test can be administered to individuals or groups in a maximum time of 10 minutes (the WEPS is untimed).

   B. Content - The 24 items consist of statements which the examinee reponds to by use of a five choice Likert Scale (strongly agree, agree, undecided, disagree, and strongly disagree).

   C. Scoring - The WEPS is scored by hand and no stencil, etc. is needed. The test administrator simply totals the weights assigned to each alternative and records its raw score number. A percentile score is given; there are several norm groups.

4. Materials Required

   A. Test sheet

   B. Pencils

   C. Manual

5. Appropriate Groups - Because the sentence structure is fairly complex and the vocabulary level of the items is high, the user will have to read at about the tenth grade level. The instrument could be used for clients who are considering jobs in organizations having a bureaucratic outlook. It could also be used as a vocationally orientated measure of a person's attitudes toward authority.

6. Technical Considerations

   A. Norm Groups - The WEPS has norms on a variety of student and worker groups of both sexes, some of these are: male and female high school and college students, Army ROTC, foremen, salesmen, mental hospital employees, and three types of public school administrators. The sample sizes appear adequate, but the samples are not described in enough detail for the reviewer to judge their representativeness.

---

*Within the context of this test, "bureaucratic" refers to a particular type of organization structure found most commonly in industry, government, the military, and education.

B. Reliability - Two types of reliability data are reported in the manual: first, internal consistency measures on four groups are: .83, .84, .89, and .91. Second, test-retest reliability studies review a correlation of .82 for a short time period and .65 after 16 months. For an instrument with only 24 items, these coefficients are acceptable.

C. Validity - The manual presents considerable validity data in the form of correlations with scales on other personality tests as well as attitude scales on authoritarianism, dogmatism, etc. Validity data based on peer ratings, performance ratings, and other criteria are given. Taken as a whole, the data presents the WEPS as being a valid instrument.

7. APA Level - A

8. Sources of Information - MMY 8:712

9. Available From - The Psychological Corporation

10. Comments - The WEPS is designed to assess one "personality trait"--the ability to work in a bureaucratic environment. This trait should be considered when assessing people for positions which commonly occur in highly structured organizations. The major problems in using the WEPS are the high reading level and the lack of enough information to judge the adequacy of the norms. However, the validation data indicate that the instrument has proven useful in many situations.

CULTURE FAIR INTELLIGENCE TEST .(IPAT OR CATTELL)

Scales 2 and 3; Forms A and B (1973)

1. Purpose - The IPAT is designed to "measure individual intelligence in a manner designed to reduce, as much as possible, the influence of verbal fluency, cultural climate, and educational level."

2. Final Score - A single final intelligence score is given both in percentile and normalized standard score I.Q.

3. Description

   A. Administration - This pencil-and-paper test can be group or individually administered in about 30 minutes. Forms A and B have identical formats and both are administered in succession. The four IPAT tests in each form are timed and examinees may mark their answers in the test booklets or on separate sheets. Scale 2 is designed for ages 8 through 13 and average adults; Scale 3 for ages 10 through 16 and superior adults. No reading is required. Optional tape recorded administrations are available for Scale 2. A Spanish edition is also available.

   B. Content - Scales 2 and 3 have four sub-tests: (1) series in which the examinee completes a sequence of four drawings by choosing one from five options, (2) classifications in which he selects one of five drawings different from the rest, (3) matrices in which he selects a drawing to complete a matrix, and (4) conditions in which he selects from among five drawings of overlapping geometric figures the one in which one or two dots could be placed to fit the specifications of a model.

   C. Scoring - It is estimated that the booklets can be hand scored in about five minutes. If separate answer sheets are used in testing, stencils are used for scoring. The raw scores for both forms are added together and these are totaled to produce a single raw score. Conversion tables are used to give the final score. Machine scoring is available for Scale 2.

4. Materials Required

   A. Test booklets

   B. Pencils

   C. Answer sheet (optional)

   D. Stopwatch

   E. Scoring stencils

5. Appropriate Groups - The Culture Fair Intelligence Tests were intended to measure the intellectual level of persons who cannot take verbal testing material. Studies have demonstrated that this test can be successfully administered to disadvantaged, illiterate, deaf, and others who are culturally, physically, or mentally different. It has been used with many non-English language groups.

6. Technical Considerations

   A. Norm Groups - Scale 2 was normed on "4,328 males and females sampled from various regions of the United States and Britain. Scale 3 norms are based on 3,140 cases, consisting of American high school students . . . and young adults in a stratified job sample." The manual does not contain an adequate description of the standardization samples or how they were selected or tested.

   B. Reliability - The manual gives "average reliabilities across samples" for each scale. While these averages are acceptable, much more information is needed before the reader can adequately judge the reliability of the tests.

   C. Validity - Validity, presented in terms of correlations with other tests of general intelligence, are quite high. However, studies predicting academic performance and school achievement yield much lower correlations coefficients.

7. APA Level - B

8. Sources of Information - MMY 8:184; MMY 6:453; MMY 5:343; MMY 4:300; Anastasi, 289-291; Cronbach, 277; Thorndike & Hagen, 666.

9. Available From - Institute for Personality and Ability Testing

10. Comments - This measure of general intelligence is very useful in the evaluation setting. Because it does not depend to any great extent on culturally loaded material, the test can be used with a number of people who would otherwise be difficult to test. The major problem is the lack of information in the manual on the technical aspects of the test making it difficult to judge the value of the test.

PEABODY PICTURE VOCABULARY TEST (PPVT)

Forms A and B (1965)

1. Purpose - The PPVT is "designed to provide an estimate of a subject's verbal intelligence through measuring his hearing vocabulary." According to the manual, it can be used as a clinical tool with "nonreaders, speech-impaired, cerebral palsied, autistic, and partially sighted" persons.

2. Final Score - The single raw score can be converted into: (1) mental age, (2) intelligence quotients, and (3) percentiles.

3. Description

   A. Administration - This untimed test is individually administered in between 10 and 15 minutes. The directions and the items are read to the examinee, who points to the correct item on each plate. The examiner reads each item aloud for the Form A or Form B answer sheet; both forms use the same set of illustrations.

   B. Content - The test contains 150 items. Each item contains four separate illustrations of people or objects. The items are arranged in order of difficulty. Not all items are administered to each examinee; the test is started at different points depending upon the age of the examinee.

   C. Scoring - Picture numbers as indicated by the subject are written on the Individual Test Record by the administrator. Incorrect responses are recorded by drawing a line through the geometric figures corresponding to the chronological number. The score is the number of correct items between the basal level (eight correct consecutive responses) and a ceiling level (six errors in eight items). All tests are scored by hand.

4. Materials Required

   A. Book of plates

   B. Test manual

   C. Individual test scoring record

   D. Pencil

5. Appropriate Groups - The PPVT is designed to be used with persons who cannot read or write. The wide range of items makes it a useful measure for testing almost the entire range of intelligence. The PPVT can be used with people who do not perform well on group tests for emotional, cultural, physical, or intellectual reasons. However, the manual states that the test is designed for persons under 18.

46

6. Technical Considerations

   A. Norm Groups - The PPVT was standardized on 4,012 cases ranging in age from two to 18 years. The subjects were white children and youth in the Nashville, Tennessee area. No adult norms are available.

   B. Reliability - Alternate form reliability coefficients range from .67 to .84. Test-retest and alternate form studies have been conducted in regular classroom students, mentally retarded persons, physically handicapped students, deaf, and emotionally disturbed students. These coefficients are all very acceptable.

   C. Validity - The manual lists summaries of numerous studies which validated the PPVT against other intelligence tests, school outcomes, etc. The subjects for these studies were usually special education and/or handicapped students. The results indicate that the PPVT has a good degree of empirical validity.

7. APA Level - B

8. Sources of Information - MMY 8:222; MMY 7:417; MMY 6:530; Anastasi, 285; Bolton, 69; Cronbach, 254.

9. Available From - American Guidance Service

10. Comments - The PPVT is designed to give a single measure of verbal intelligence in a short period of time. The test is well constructed and the validity data is very encouraging. Because of its format and administration procedures, it can be used with a wide range of disabilities. From a rehabilitation point of view, the major limitation of the PPVT is the lack of adult norms.

REVISED BETA EXAMINATION

Second Edition - Beta II (1978)

1. Purpose - The Beta II is "designed to measure the general intellectual ability of persons who are relatively illiterate, non-English speaking, or suspected of having other language difficulties. It can also be used as a nonverbal measure for members of the general population." It must be pointed out that the Beta II is a complete and thorough revision of the previous editions of the Revised Beta Examination.

2. Final Score - The Beta II yields a single intelligence score which can be presented either as an I.Q. or a percentile score.

3. Description

   A. Administration - The test can be either group or individually administered; group administration takes about 30 minutes. Prior to each of the six sub-tests, the test administrator reads the instructions and monitors the completion of the practice exercises. All answers are recorded in the test booklet; there is no second answer sheet. Spanish instructions are contained in the manual.

   B. Content - The Beta II contains six sub-tests: (1) mazes - 5 items; (2) coding - 90 items; (3) paper form boards - 18 items; (4) picture completion - 20 items; (5) clerical checking - 56 items; and (6) picture absurdities - 21 items. All tests are highly speeded.

   C. Scoring - The Beta II is hand scored with a stencil. A separate raw score is obtained for each sub-test. After the raw scores are converted to weighted scores, the six weighted scores are added to give a total score, which is then converted into a single I.Q. and/or percentile score. Scoring should take about five minutes.

4. Material Required

   A. Test booklet

   B. Pencils

   C. Administration manual

   D. Scoring stencil

   E. Enlargement of paper form board practice problems

5. Appropriate Groups - The test appears to be very appropriate for the groups for which it was designed--persons who are nonliterate or semiliterate in written English. The test does not require reading, however, the examinee must be able to hold a pencil, have enough dexterity to trace mazes, fill in missing parts on pictures, print numbers, and make "X's." The Spanish directions also increase the potential usefulness and range of the test. The Beta II is intended for use with a general adult group. "However, the ceiling of the examination is not high, and it

does not differentiate as well among excpetionally able examinees as it does among those of lower ability."

6. Technical Considerations

   A. Norm Groups - The test was standardized on a national stratified sample of 1,050 persons between the ages of 16 and 64. Sex, racial composition, occupation, and geographic region were considered. The sampling plan was well thought out and all samples are described in detail. Norms are available on seven age groups: (1) 16-17, (2) 18-19, (3) 20-24, (4) 25-34, (5) 35-44, (6) 45-54, and (7) 55-64.

   B. Reliability - A test-retest correlation of .84 over a three week interval was found. While this is adequate, the Beta II manual should contain additional reliability data.

   C. Validity - The manual attempts to establish the validity of the Beta II in two ways. First, correlations between the first Beta edition and the Beta II were established. Because these were high (.84 and .93), it is assumed that validity studies established for the first edition would apply to the Beta II. This indirect process of validation must be replaced by more direct procedures. Second, correlations between Beta II and WAIS I.Q.'s are given for two age groups--18 to 19 and 35 to 44. Full scale I.Q.'s correlated .64 and .66 with the Beta II.

7. APA Level - A

8. Sources of Information - (earlier editions) MMY 6:494; MMY 3:259.

9. Available From - The Psychological Corporation

10. Comments - This complete revision of the Beta resulted in a general nonverbal intelligence test that can be widely used in evaluation. The norm base is current and adequate and test administration procedure considers that persons with little (positive) experience with testing will be given the Beta II. The Beta II has good potential if the reliability and validity are more firmly established through further research.

SRA PICTORIAL REASONING TEST (PRT)

Self-Scoring Edition (1973)

1. Purpose - The PRT is a test of "General ability . . . developed empirically to minimize differences across major American cultural subgroups." It is used to "measure the learning potential of individuals from diverse backgrounds with reading difficulties, whose potential for training and employment cannot be reliably and validly measured by verbal instruments."

2. Final Score - A single final score is given in percentiles.

3. Description

   A. Administration - The PRT is a group administered pencil-and-paper test in which the administrator reads the directions to the examinees. The test contains ten practice items and gives the reasons why each item should be marked as it is. The test can be given either as timed (15 minutes) or untimed. Because the test is designed to provide "maximum cultural fairness" with the untimed administration, this form of administration is recommended. It is estimated by the reviewer that the untimed administration would take about 25 minutes. During testing the examinee marks his answer in a self-scoring carbon-centered booklet.

   B. Content - In this 80 item test, each item is made up of five pictures, four of which are related in some way. Typical items contain drawings of objects or designs. In some of the items it is difficult to discriminate some of the details of the drawings.

   C. Scoring - The examiner counts the number of correct responses, totals them, and compares the single raw score to the appropriate norms table to obtain a percentile score.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Manual

   D. Stopwatch (timed condition only)

5. Appropriate Groups - The PRT can be used with persons 14 years old or older who are able to see well enough to perceive the items. The test is intended to be as culture-fair as possible and is intended for use with persons who have a high school education or less--"many test items are ambiguous for those with a higher educational background in that such individuals tend to read too much into an item and thus draw the wrong conclusions." The PRT is best used with persons of low educational attainment who are culturally different from middle class culture.

6. Technical Considerations

   A. Norm Groups - A variety of timed and untimed norms are available on groups such as: students in grades 9 through 12; white and nonwhite employed persons in office positions, craft jobs, and a variety of semi-skilled jobs. While sample sizes are adequate, samples are only partially described. The user needs additional information on the geographical distribution of the samples.

   B. Reliability - Internal consistency (KR-20) reliabilities are the only estimates of reliability given in the manual. These range from .59 to .83 and are somewhat lower than what is expected in a test of this type.

   C. Validity - The PRT correlates moderately well with the WAIS and the Lorge-Thorndike Intelligence Tests. Correlations between PRT and school grades are generally low, with most being non-significant. Finally, correlations with an overall rating of job success on 25 separate occupations produced mixed results--there were more significant correlations in semi-skilled jobs than in jobs that required greater degrees of skill. These attempts at validation lead to the conclusion that the test has already demonstrated some degree of validity.

7. APA Level - A

8. Sources of Information - MMY 7:381

9. Available From - Science Research Associates

10. Comments - the PRT can be used to provide a general indication of how well a client functions--especially a nonreading client or a client from a different culture. The test is intended to eliminate or reduce cultural bias, and according to data published in the manual, it partially achieves this goal. The major technical problem with the test is the fairly low reliabilities. Some of the objects in the drawings are not clear--this presents another problem. In conclusion, this test can be best used when the evaluator needs to know what the intellectual level of functioning of a nonliterate person with normal vision.

SRA VERBAL FORM

Form A (1973)

1. Purpose - "The SRA Verbal is a test of general ability. It is used as a measure of an individual's overall adaptability and flexibility in comprehending and following instructions, and in adjusting to alternating types of problems."

2. Final Score - The test gives three final scores: (1) L or Linguistic, (2) Q or Quantitative, and (3) Total Score. The Total Score is a combination of the L and Q scores. Scores are given in percentiles. The manual warns that the Q and L scores should not be used separately.

3. Description

   A. Administration - In this group administered paper-and-pencil test, the examinee marks his answers on a self-scoring carbon-centered booklet. The SRA Verbal Form is a speeded test--15 minutes are allowed to answer 84 items. The manual contains no estimate of the reading level required to take this test; however, the reviewer estimates that an eighth grade reading ability is required.

   B. Content - The test contains 84, five-alternative multiple-choice items arranged by increasing difficulty in the following sequence: two same or opposite items, one mathematics problem solving, two word definitions in sentences, and two number series.

   C. Scoring - The examiner counts the number of correct L and Q responses and then totals them. If the educational norms are used, separate percentiles are available for L, Q, and Total. In using industrial norms, the Total Score is the only score that is converted.

4. Materials Required

   A. Test booklet

   B. Pencil

   C. Manual

   D. Stopwatch

5. Appropriate Groups - This test could be used in assessing a person's general level of functioning. As opposed to nonverbal measures, the SRA Verbal Form contains material which is closely related to general intelligence and academic functioning. The SRA Verbal Form is useful with clients who are considering additional education or jobs that require a high degree of general intellectual functioning.

6. Technical Considerations

   A. Norm Groups - There are two separate types of norms. The first are educational norms on students from ages 12 through 17; this reviewer was unable to find any information about this sample in the manual. The second are employed worker norms on 13 specific occupations ranging from retail store managers to unskilled laborers. Some minority group norms are available; a few of these occupations were almost 100% women dominated. These sources of the subjects are adequately explained and many of the sample sizes are of adequate size.

B. Reliability - Equivalent form reliability between forms A and B was established on 300 high school students. Correlations on the three scales were: L = .76; Q = .80; and Total = .78. While these are adequate, the reviewer would like to see internal consistency reliability estimates.

C. Validity - The manual presents correlations with many other intelligence or general ability tests (e.g., Otis Test of Mental Ability; Army General Classification Test) which range from .18 to .82, with most correlations being in the .60's. Correlations using concurrent designs between the test and overall rating of employed workers were mixed--about half showing significant results. However, one well controlled study reported significant findings for several age and racial groups of clerical workers. In conclusion, the results given in the manual are promising.

7. APA Level - A

8. Sources of Information - (earlier edition) MMY 7:383; MMY 4:319.

9. Available From - Science Research Associates

10. Comments - This test is intended to be used in making hiring and other placement decisions where a measure of general ability is required for job performance or educational planning. The industrial norms are useful to the evaluator who knows not to confuse norms with validity. The validity studies show the potential of the SRA Verbal Form. However, the test is not culture fair or culture free and was never intended to be. It is designed to be used in settings where the traditional academic standards apply--in the job market or in education.

DIFFERENTIAL APTITUDE TEST (DAT)

Forms S and T (1974)

1. Purpose - Originally developed in 1947, these new forms of DAT retain the original purpose of the instrument--"to provide an integrated, scientific, and well standardized procedure for measuring the abilities of boys and girls in grades eight through twelve for purposes of educational and vocational guidance." These tests have also been used with young adults and in employee selection.

2. Final Score - The battery yields nine aptitude scores: (1) verbal reasoning, (2) numerical ability, (3) abstract reasoning, (4) clerical speed and accuracy, (5) mechanical reasoning, (6) space relations, (7) spelling, (8) language usage, and (9) a general mental ability score which is the combination of the first two scores. Scores are presented in percentiles and stanines.

3. Description

   A. Administration - This is a group administered pencil-and-paper test. Total testing time takes about four hours, which can be broken into: (a) two 2-hour sessions, (b) four 1-hour sessions, or (c) six 35 to 45 minute sessions. All tests are timed, but only the clerical speed and accuracy sub-test can be regarded as a highly speeded test. The eight tests are available in two reusable formats: a one booklet edition (Forms S and T) and a separate booklet for each test (Form T). Both formats require the use of a separate answer sheet of which five are available: MRC, OpScan, NCS, IBM 805, and IBM 1230.

   B. Content - While the tests contain a variety of different materials, all items except the spelling sub-test are five alternative multiple-choice. The following item types are used (the numbers in parenthesis correspond to the number in the "Final Score" heading above): (1) double-ended analogies - 50 items; (2) arithmetic computation - 40 items; (3) progression of forms - 50 items; (4) two letter or number combinations - 200 items; (5) illustrations of machinery to illustrate various physical laws - 70 items; (6) patterns to be "mentally folded" into figures - 60 items; (7) words spelled correctly or incorrectly - 100 items; and (8) detecting errors in grammer, punctuation, and capitalization - 60 items.

   C. Scoring - Tests can be either hand or machine scored. Hand scoring requires the use of stencils and is estimated to be 15 minutes. Scores are plotted on a report form for easy interpretations. Each score is presented in terms of "score band"--a standard error of the measurement.

4. Materials Required

   A. Test booklet(s)

   B. Answer sheet

   C. Pencils

D. Manual

E. Stencils

F. Stopwatch

G. Report form

5. Appropriate Groups - The DAT is aimed at young persons who are planning their vocational futures. In vocational rehabilitation the DAT is useful when an overall measure of the person's aptitudes is needed to assist in making vocational and educational plans. The length of the battery, the need for about a sixth grade reading level, and the use of a separate answer sheet restrict the use of the DAT to fairly high functioning persons.

6. Technical Considerations

   A. Norm Groups - Norms were derived from over 63,000 school students in the eighth through the twelfth grades. Separate norms are available by sex and grade. The sample was carefully selected to represent the public school population of the United States. The sample is clearly described.

   B. Reliability - Split-half reliabilities were computed for all subtests except the clerical speed and accuracy. Because of its speeded conditions, test-retest reliabilities were used. These estimates of reliability are given for each norm group. Most of the split-half and test-retest coefficients are in the low .90's. This is one indication of the careful construction of the battery.

   C. Validity - The DAT has been thoroughly and successfully validated against a variety of secondary school grades, achievement tests, and aptitude tests. The evidence given against these criteria is impressive. A problem in the validation of the DAT is a lack of studies using job or taining success as the criterion. The test is useful in predicting academic outcomes and, in spite of a lack of job related studies, it is one of the best researched batteries available.

7. APA Level - B

8. Sources of Information - MMY 8:485; (earlier editions) MMY 7:673; MMY 6:766; MMY 5:605; MMY 4:711; MMY 3:620; Anastasi, 130-131, 147-149, 379-383; Barnette, 154-159; Bolton, 96; Cronbach, 353-355, 359, 362, 368; Guion, 264-266; Horrocks & Schoonover, 386-392; Lyman, 140-141; Noll & Scannell, 271-273; Super & Crites, 339-349; Thorndike & Hagen, 236-237, 350-353, 358-360, 528; Tyler, 62-63.

9. Available From - The Psychological Corporation

10. Comments - The DAT is an excellently constructed and thoroughly researched battery. It would especially be useful in providing general measures to be used in planning education and training. While it is impossible not to recommend it on its technical qualifications, the use of this battery in vocational evaluation may present problems because of: (1) length, (2) reading skills required, (3) use of separate answer sheets, and (4) the abstract nature of some sub-tests.

USES GENERAL APTITUDE TEST BATTERY (GATB)

B-1002, Forms A and B (1970)

1. Purpose - The GATB is intended to assess "vocationally significant aptitudes" that are useful in vocational counseling, training selection, job selection, and job placement.

2. Final Score - The GATB is composed of 12 sub-tests (parts) which yield measures of nine aptitudes: (1) general learning ability, (2) verbal, (3) numerical, (4) spatial perception, (5) form perception, (6) clerical perception, (7) motor coordination, (8) finger dexterity, and (9) manual dexterity. Aptitude scores are presented in standard scores with a mean of 100 and a standard deviation of 20.

3. Description

   A. Administration - The GATB is a timed group test which takes about two and one half hours to administer. Seven of the sub-tests are printed in two booklets, one on a single sheet, and four sub-tests involve the use of apparatus. Separate answer sheets (three types are available: NCS, IBM 1230, and OpScan) are required and the examiners must stand for two of the apparatus parts. Sturdy tables about 30 inches in height are required for the apparatus tests.

   B. Content - The pencil-and-paper tests contain a wide variety of item types and formats: (1) matching names - 150 items, (2) arithmetic computation - 50 items, (3) patterns to be "mentally folded" into figures - 40 items, (4) vocabulary - same-opposite - 60 items, (5) matching tools - 49 items, (6) arithmetic reasoning - 25 items, (7) form matching - 60 items, and (8) rapid pencil movement - 130 spaces to be filled. The apparatus tests use a 48 hole pegboard and a 50 hole rivet board.

   C. Scoring - Scoring may be performed by hand or machine. Hand scoring takes about 20 minutes. Using an Individual Aptitude Profile Card, scores are converted and some are combined to form the nine aptitudes. In an attempt to compensate for possible misinterpretation, one standard error of the measurement is added to each aptitude score. Aptitude scores are compared to requirements for specific jobs (Specific Aptitude Test Batteries - SATB's) or a group of jobs having the same aptitudes (Occupational Aptitude Patterns - OAP's). This comparison can be done by hand or machine.

4. Materials Required

   A. Booklets I and II

   B. Answer sheet

   C. Sheet for Part 8 (motor coordination)

   D. Pegboard

E. Finger board

F. Pencils

G. Administration and scoring manual

H. Two norm manuals (one for OAP's and one for specific jobs)

I. Recording sheets for apparatus tests

J. Stopwatch

K. Results forms

L. Scoring keys

M. Individual Aptitude Profile Card

5. Appropriate Groups - The GATB can be used with a wide range of groups. Research has indicated that it can be successfully given to mentally retarded, deaf, and emotionally disturbed clients. It has also been successfully used with many English speaking minority groups. (A Spanish language version is also available.) The sixth grade reading level. however, makes the test inappropriate for disadvantaged persons and persons with low reading skills. Because examinees are required to stand for the pegboard parts, wheelchair persons may have a difficult time. The fine dexterity necessary for the rivet board prevents its use with many persons having upper extremity handicaps.

6. Technical Considerations

   A. Norm Groups - The GATB was originally normed on 4,000 workers which were a random sample of the "general working population." Since that time additional norms have been developed for ninth and tenth grade students. All samples are clearly and accurately defined.

   B. Reliability - The development section of the manual contains a chapter on reliability, and these extensive results cannot be summarized in a few lines. The reliabilities for the pencil-and-paper tests are very high, and the apparatus tests yield moderately high reliability coefficients.

   C. Validity - The strong point of the GATB is its validity. The battery has been validated on over 600 jobs. Extensive data relating the GATB to training success, educational attainment, and other tests are available. The employment service's ongoing program of test validation will hopefully make certain that the GATB continues to be the best validated aptitude test available.

7. APA Level - B

8. Sources of Information - MMY 8:490; MMY 7:676; MMY 6:771; MMY 4:714; Anastasi, 184-185, 384-387; Barnette, 117-129, 154-159; Bolton, 189-198; Cronbach, 355-363, 366; Guion, 269; Super & Crites, 330-339; Thorndike & Hagen, 353-357, 360-366, 528, 610-611; Tyler, 58-60.

9. Available From - Persons interested in this battery must contact their State Employment Service. (Do not contact the U.S. Department of Labor.) The GATB is released to nonprofit organizations engaged in counseling. Formal training is required prior to release.

10. Comments - Although the GATB is a widely used and technically superior battery, it has two problems: (1) much of the validation research has been on semiskilled occupations and not on skilled or technical jobs and (2) the multiple-hurdle method of validation developed prior to computer technology could be replaced by a multiple regression procedure. The major advantages are (1) the relationship of the GATB to jobs and training and (2) the relationship between aptitude scores and the classification systems used in the Dictionary of Occupational Titles.

USES NONREADING APTITUDE TEST BATTERY (NATB)

Form A (1971)

1. Purpose - The NATB is a nonreading version of the General Aptitude Test Battery (GATB) and measures the same nine aptitudes as does the GATB. The NATB is intended for use with disadvantaged individuals who lack the literacy and test-taking skills necessary to complete the GATB.

2. Final Score - The NATB contains 14 tests which yield measures of the same nine aptitudes as the GATB (page 51). Aptitude scores are presented in standard scores with a mean of 100 and a standard deviation of 20.

3. Description

   A. Administration - The NATB can be administered to groups of up to six at a time. The first nine tests are contained in eight separate booklets. The other five tests are identical to the GATB motor coordination and dexterity tests. The administrator reads the items aloud for the two vocabulary tests. The numerical aptitude is assessed in part by coin matching and coin series tests. Other tests are modified versions of the GATB matching names, patterns, matching tools, and form matching. Enlargements of the pages in the test booklets are used to help the examinees locate and mark the answers. All answers are marked in test booklets--there are no separate answer sheets. Total testing time is estimated to be over three hours.

   B. Content - The pencil-and-paper booklets contain items similar or identical to the GATB, with two major exceptions: (1) there are no arithmetic problems (numerical aptitude is measured by the mental manipulation of coins), and (2) vocabulary items are presented orally.

   C. Scoring - Scoring may be done by hand or machine. Hand scoring requires about 20 minutes. Scores are converted and OAP norms are used. The procedure is exactly the same as the GATB.

4. Materials Required

   A. Books 1 through 8

   B. GATB Part 8 (motor coordination)

   C. Pegboard

   D. Finger board

   E. Pencils

   F. Administration and scoring manual

   G. The two GATB norms manuals

H. Recording sheets for apparatus tests

I. Enlargements of pages from test books

J. Stopwatch

K. Results forms

L. Scoring keys

5. Appropriate Groups - The NATB was specially designed for disadvantaged persons in an attempt to eliminate the barriers of language and test sophistication. Because the NATB was designed for disadvantaged, it should not be used with persons who function extremely well on tests--the test results will not be accurate. Although the manual contains no information, the battery appears to be potentially useful for assessing the overall vocational potential of the mentally retarded. Physically handicapped persons are limited on this battery in the same ways they are on the GATB.

6. Technical Considerations

   A. Norm Groups - The NATB was standardized on 848 high school students. The sample characteristics are clearly described. Regression equations were used to equate NATB to GATB scores.

   B. Reliability - The manual contains no reliability data as such. Because several of the tests were modified from GATB tests, they are assumed, by the developer, to be reliable. This lack of reliability data is a definite weakness of the manual. However, one estimate of stability, the standard error of the measurement, is given for each aptitude; these appear to be within reason.

   C. Validity - The NATB bases its validity on its correlations with the GATB. Validation studies are presently underway, none have been published in a revised manual.

7. APA Level - B

8. Sources of Information - MMY 8:491; MMY 7:679; Anastasi, 387; Bolton, 198-202.

9. Available From - Persons interested in this battery must contact their State Employment Service (do not contact the U.S. Department of Labor). The NATB is released to nonprofit organizations engaged in counseling. The NATB requires a two and one-half day training session prior to use.

10. Comments - The major use of the NATB is assessing a wide range of aptitudes that have been proved to be vocationally significant. The fact that NATB is nonreading makes it very useful to a wide variety of client problems. As with the GATB, scores can be related to the Dictionary of Occupational Titles. Because the NATB lacks the technical rigor of its parent, the GATB, the evaluator "must be extremely cautious and conservative in interpreting scores on the NATB." Despite the claims of its manual, the NATB should not be used for selection or placement because of its lack of direct job validation.

GENERAL CLERICAL TEST (GCT)

(1972)

1. Purpose - The GCT "has been designed to measure aptitudes which are of importance in clerical work of all kinds." The GCT has been used for selection and placement in education and industry.

2. Final Score - There are four final scores: (1) clerical, (2) numerical, (3) verbal, and (4) total, which is the sum of the first three. All scores are given in percentiles and several norms are available.

3. Description

   A. Administration - This is a group administered pencil-and-paper test in which all answers are marked or written in the test booklet. There are a total of ten sub-tests; all are highly speeded. The reviewer estimates that total administration time to be from 50 to 55 minutes. The test is published in two formats; a 12 page booklet and two 6 page booklets.

   B. Content - The GCT contains ten sub-tests: (1) checking - comparing names, addresses, and amounts and picking out the differences - 19 items; (2) alphabetizing - recording the number of a file drawer for names - 61 items; (3) arithmetic computation - 20 items; (4) error location - using addition and subtraction to find the error in a matrix - 20 items; (5) arithmetic reasoning - reading problems - 16 problems; (6) spelling - finding and correctly spelling words - 29 items; (7) reading comprehension - answering questions about two paragraphs - 14 items; (8) vocabulary - selecting the word that means the same - 40 items; and (9) grammar - finding and correcting the error in a sentence - 24 items. The clerical score is totaled from tests 1 and 2; the numerical from 3, 4, and 5; and the verbal from 6, 7, 8, and 9. Only tests 7 and 8 are multiple-choice; the rest require a written answer.

   C. Scoring - All scoring is done by hand with a folded answer sheet.

4. Materials Required

   A. Test booklet (or booklets)

   B. Pencils

   C. Manual

   D. Stopwatch

   E. Answer sheet

5. Appropriate Groups - The purpose of the GCT is to predict employment or training success. The test can be used in a vocational evaluation situation to assess an overall ability for clerical work. Obviously, to successfully take this test, the client needs good visual acuity, use of at least one hand, and an estimated sixth grade reading level.

6. Technical Considerations

   A. Norm Groups - Norms are available on female students in high schools and private business schools, for numerous applicant groups, and several groups of persons employed in clerical positions. Most norm groups are of adequate size with most of the norms developed in the late 1960's and early 1970's. The group characteristics are not described as thoroughly as they should be.

   B. Reliability - The results of two test-retest studies are given in the manual; the average correlation is in the low .90's. Standard error of measurement results are also given. Both sets of statistics imply that the GCT has adequate reliability for its intended use.

   C. Validity - The manual contains two types of validity evidence. First are correlations between test results and school grades and grade point averages. These are yielded significant results. Second are studies in which the GCT was correlated with job success criteria (usually supervisory ratings). The results of these are mixed. The test appears to have some degree of validity.

7. APA Level - A

8. Sources of Information - MMY 8:1032; (earlier editions) MMY 4:730; MMY 3:630; Guion, 242-243.

9. Available From - The Psychological Corporation

10. Comments - The GCT has been renormed and updated since the original 1944 edition, thus giving this old, established test a new life. The major use of the GCT would be to provide an overall assessment of the client prior to in-depth assessment for specific clerical jobs.

MINNESOTA CLERICAL TEST (MCT)

(1959)

1. Purpose - The MCT is a "test of speed and accuracy in performing tasks related to clerical work." The test assesses the ability to rapidly perceive and differentiate between numbers, letters, and other symbols.

2. Final Score - There are two final scores: (1) number comparison and (2) name comparison. Several norm groups are available; each score is reported in percentiles.

3. Description

   A. Administration - This highly speeded group administered pencil-and-paper test takes about 15 minutes to administer--8 minutes for the numbers and 7 minutes for the names. For each sub-test, the examinee compares two names or numbers to see if they are identical. Answers are marked in the test booklet; there is no separate answer sheet.

   B. Content - The test consists of two parts (number checking and name checking) each having 200 items. For each sub-test the examinee compares two numbers or names to see if they are identical or not.

   C. Scoring - The instrument is hand scored in about five minutes using stencils. Raw scores on each part are converted to percentiles using the appropriate norm group.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Manual

   D. Stopwatch

   E. Scoring stencils

5. Appropriate Groups - Because the test items are designed for perception per se and not reading, in theory, reading level should not be an important consideration. However, clients who read well have a definite advantage. Good vision is required. The examinee should be able to relate the content to clerical duties; face validity is high.

6. Technical Considerations

   A. Norm Groups - Separate norms are available by sex for grades 8 through 12 and on male and female employed workers in a variety of clerical positions. Although the norm groups for the most part are well described, some norms are over 30 years old.

   B. Reliability - Test-retest reliabilities are reported for two groups of students and three groups of employed workers. Correlations range between .56 and .93, with most values being in the .60's and .70's.

C. Validity - There are several studies listed in the manual which correlate MCT scores with various criteria of job performance. In general, these correlations are high. Correlations with other clerical tests are also listed.

7. APA Level - A

8. Sources of Information - MMY 8:1036; MMY 6:1040; MMY 6:850; MMY 3:627; Anastasi, 449-451; Bolton, 5, 8-9; Guion, 240-242; Super & Crites, 162-179.

9. Available From - The Psychological Corporation

10. Comments - The MCT is an old, widely used, respected clerical test. Evaluators who assess client clerical ability with this device should recall that many of the norms are most likely out-of-date. Job requirements and hiring standards have changed in the 30 years since this test was originally normed.

SRA TYPING SKILLS

Forms A and B (1947)

1. Purpose - The typing test is designed to measure typing achievement.

2. Final Score - There are two final scores: (1) net speed and (2) accuracy. These are converted into percentiles.

3. Description

   A. Administration - The test is group administered on manual or electric typewriters. The examiner gives each examinee a sheet containing instructions and the material to be typed. The examinee reads the instructions and then practices by typing the instructions. The test consists of typing a business letter as many times as possible during a ten minute period. The letter is typed on a special work sheet.

   B. Content - A four paragraph, 225 word business letter with a salutation and complimentary closing are typed from a clear printed copy.

   C. Scoring - The examiner first finds the total number of strokes and the total number of errors. Using graphs, he converts these to raw scores which are compared to norms.

4. Materials Required

   A. Typewriter

   B. Test form

   C. Scratch paper

   D. Manual and norms sheets

   E. Stopwatch

5. Appropriate Groups - The test is appropriate for individuals who have had some experience in typing and are considering office careers.

6. Technical Considerations

   A. Norm Groups - Norms are available on inexperienced and experienced office job applicants using manual and electric typewriters. Net speed norms are available for students. The sample characteristics of the groups are not described in nearly enough detail; nor are the applicant norms based on a random sample. Apparently, the test was renormed in the late 1960's and early 1970's.

   B. Reliability - No reliability studies are reported in the manual.

   C. Validity - No validation studies are reported in the manual.

7. APA Level - A

8. Sources of Information - MMY 8:1036; MMY 6:51; Guion, 419; Super & Crites, 156-157.

9. Available From - Science Research Associates

10. Comments - The test has three major problems: (1) the content of the letter is more representative of straight typing than of a business letter, (2) unknown reliability, and (3) unknown validity. This test should be used in conjunction with other typing tests and work samples.

STENOGRAPHIC APTITUDE TEST (BENNETT)

(1939)

1. Purpose - "This test is designed to predict ability to acquire the skills of shorthand and typewriting." The test is "valuable in determining whether or not an individual is likely to profit by secretarial training."

2. Final Score - The test yields three scores: (1) transcription, (2) spelling, and (3) total, a combination of the transcription and spelling scores. Scores are reported in percentiles.

3. Description

   A. Administration - This pencil-and-paper test can be administered to a group in less than 25 minutes. Both the transcription and spelling parts are timed, and both require the examinee to write the answers in the test booklet. Unlike most clerical tests, this does not appear to be a highly speeded test.

   B. Content - In the transcription part, the examinee is given five minutes to substitute symbols for the numbers one through five. There are 22 lines of ten numbers each. The examinee then transcribes the symbols back to numbers. This part is not timed. The spelling sub-test requires the examinee to determine if 100 words are correctly or incorrectly spelled. If incorrectly spelled, the examinee spells the word correctly.

   C. Scoring - The test can be hand scored in about two minutes. Raw scores are converted to percentile scores by using the appropriate norm groups.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Manual

   D. Stopwatch

   E. Scoring key for spelling

5. Appropriate Groups - The test is appropriate for clients who are considering training or a career in the secretarial field. Due to the folding of the test form for the transcription part, left-handed clients may have some difficulty taking the test.

6. Technical Considerations

   A. Norm Groups - Percentile norms are based on 870 high school girls with no shorthand experience and 345 girls with less than one semester of shorthand experience. Supplementary norms for private stenographic school students are given. Samples are not described, and the norms were developed about 30 years ago.

B. Reliability - Even-odd reliability for the transcription test is reported as being .975; split-half reliability for the spelling test as .913. The manual does not contain enough information to adequately interpret these coefficients.

C. Validity - While the manual reports some correlations with training outcomes, the lack of information makes these results almost meaningless.

7. APA Level - A

8. Sources of Information - MMY 3:390.

9. Available From - The Psychological Corporation

10. Comments - The Stenographic Aptitude Test is a practical device intended to predict training success. Because of technical problems with norms, reliability, and validity, the evaluator should use this in conjunction with other measures of clerical ability.

AAMD-BECKER READING-FREE VOCATIONAL INTEREST INVENTORY (R-FVII)

Experimental Version (1975)

1. Purpose - "The Inventory was devised to provide systematic information on the interest patterns of mentally retarded males and females engaged in occupations at the unskilled and semiskilled levels." This vocational perference test helps to identify areas in which persons have occupational interests.

2. Final Score - The R-FVII has separate male and female forms. The male form contains the following areas: (1) automotive, (2) building trades, (3) clerical, (4) animal care, (5) food service, (6) patient care, (7) horticulture, (8) janitorial, (9) personal service, (10) laundry service, and (11) materials handling. The female form yields eight final scores: (1) laundry service, (2) light industrial, (3) clerical, (4) personal service, (5) food service, (6) patient care, (7) horticulture, and (8) housekeeping. The results are converted to both T scores and percentiles and then plotted on a profile.

3. Description

   A. Administration - The R-FVII has no time limit and can be administered in about 45 minutes to either individuals or groups. Both the male and female forms can be administered at the same time. The examinee is asked to read the instructions on the cover of the test booklet while the examiner reads them. This is the only reading required. The test is printed in a disposable booklet; answers are marked in the test booklet.

   B. Content - Each item consists of three illustrations of a person working; the examinee circles the picture of the activity he/she would most like to do. There are 40 female and 55 male items. The illustrations are simple line drawings. All illustrations in the male form show only males; the female form only females.

   C. Scoring - The R-FVII is hand scored using a grid on the last page of the test booklet; the examiner records the responses from the test booklet, totals them, and converts the raw scores to T scores and percentiles.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Administration manual

5. Appropriate Groups - According to the manual, the R-FVII is to be used "with mentally retarded persons, particularly the educable mentally retarded at the high school level." Because the test items require no reading, it could easily be used with retarded persons who do not speak English. Since the occupations shown in the pictures are of semiskilled jobs, at the most, the R-FVII should not be used for persons who could be trained for skilled positions.

6. Technical Considerations

   A. Norm Groups - Norms are available on educable mentally retarded males and females in grades 9 through 12 in public day schools. Male and female norms are also available on students in "ungraded residential institutions in the United States." The norm groups are of adequate size, but this reviewer would like more information on the geographic distribution of the samples.

   B. Reliability - The manual contains two different types of reliability data. The first are test-retest correlations over a two week interval; these range mostly in the .80's. The second are internal consistency (KR-20) coefficients, which according to the manual range from .68 to .96. The manual does not explain the methodology nor give the results in enough detail to permit the reader to accurately interpret the KR-20 estimates.

   C. Validity - The R-FVII bases its validity mainly on correlations between two other picture interest inventories--the Geist Picture Interest Inventory and the Picture Interest Inventory. Because of their lack of adequate development, neither of these provides a meaningful criterion for the R-FVII.

7. APA Level - B

8. Sources of Information - MMY 8:988.

9. Available From - American Association on Mental Deficiency

10. Comments - The test provides the user with one method of determining the vocational interests of a special group of clients. While one could argue that the choice of occupational areas and possibly even the use of a separate form for each sex is stereotyping, the scope of the items is within the range of realistic job opportunities for retarded persons. The major problems are the almost complete lack of validity and the lack of information about reliability. The R-FVII should be used with caution.

KUDER OCCUPATIONAL INTEREST SURVEY (KOIS)

Form DD (1974)

1. Purpose - The KOIS provides information that aids in making a vocational choice or selecting a tentative field of study by identifying interests in relation to occupations or occupational fields. It can be used with high school and college students as well as with adults in employment counseling and retraining.

2. Final Score - The KOIS contains a wide range of scales for both occupations and college majors. The total 171 scales contain: (1) 77 occupational scores normed on men (e.g., carpenter, machinist, and statistician); (2) 29 college major scores normed on men (e.g., biological sciences and psychology); (3) 37 occupational scales normed on women (e.g., bank clerk, lawyer, and X-ray technician); (4) 19 college major scores normed on women (e.g., English and nursing); and (5) eight experimental scales. Both men's and women's scales are used for each examinee regardless of sex. Scale scores are reported in terms of the correlations between the individual's responses and those of each group used in the development of a specific scale.

3. Description

   A. Administration - This untimed pencil-and-paper instrument is group administered in approximately 30 to 40 minutes. The items are printed on the answer sheet; no separate test booklet is required.

   B. Content - The 100 triad items each contain three short statements of different activities. The examinee chooses the one activity in each triad he prefers most and the one activity in each triad he likes least.

   C. Scoring - The KOIS must be machine scored. Each examinee receives a profile sheet plotting his results.

4. Materials Required

   A. Answer sheet

   B. Pencils

   C. Manual

5. Appropriate Groups - Persons must be able to read at the sixth grade level to use the KOIS. Because of the size of the print and the lack of contrast between print and paper colors, persons with even mild visual problems may have difficulty reading the items. There are scales for many occupations that do not require college or technical training, thus, making the KOIS useful for persons who do not desire additional formal education. However, the college major scales make it most appropriate for high school students and others who are considering formal academic training.

6. Technical Considerations

   A. Norm Groups - Each scale was developed on a separate group of employed workers or students. The scales for all occupational groups were developed on workers employed in these occupations. Students majoring in specific academic areas provided the data for the college major scales. All groups are clearly described in the manual.

   B. Reliability - Test-retest reliabilities over a two-week period are reported as .93 and .96. The reliability was also defined in terms of consistency of the differences between scores or each pair of scales. These are at acceptable levels.

   C. Validity - The manual presents no evidence that the KOIS can predict future job success based on interest. However, data on classification of presently employed workers according to their interests are presented.

7. APA Level - B

8. Sources of Information - MMY 8:1010; MMY 7:1025; Anastasi, 538-539

9. Available From - Science Research Associates

10. Comments - Because the KOIS covers a wide variety of occupations, scores on college major interest, and combines male and female occupational interests, it has a wide variety of uses within the evaluation setting.

MINNESOTA IMPORTANCE QUESTIONNAIRE (MIQ)

1967 Revision (1967)

1. Purpose - The MIQ is designed to assess an individual according to 20 vocationally relevant personal need dimensions in order to identify jobs with which he might be most satisfied.

2. Final Score - Scores are provided which represent a measure of the degree of importance assigned by the individual to each of the 20 need dimensions: (1) ability utilization; (2) achievement; (3) activity; (4) advancement; (5) authority; (6) company policies and practices (7) compensation; (8) co-workers; (9) creativity; (10) independence; (11) moral values; (12) recognition, (13) responsibility; (14) security; (15) social service; (16) social status; (17) supervision - human relations; (18) supervision - technical; (19) variety; and (20) working conditions. These scores are presented as adjusted scale values and plotted on a profile. Additional scores are provided which are measures of the degree of correspondence between the MIQ profile of the individual and similar profiles of the work reinforcer opportunities of 148 occupations.

3. Description

   A. Administration - This untimed pencil-and-paper test can be individually or group administered in less than 40 minutes. The MIQ requires the use of a separate answer sheet; test booklets are reusable.

   B. Content - Each of the 20 need dimensions is represented by a single statement (e.g., "I could be busy all the time" or "I could work alone on the job"). In the first 190 items each of the 20 statements is paired with all other statements. The individual is asked to indicate the statement of each pair which represents the more important characteristic of his ideal job. In the last 20 items each statement appears independently. The individual is asked to indicate whether each need dimension is important or not in his ideal job.

   C. Scoring - Although the device may be hand scored, machine scoring is recommended. MIQ results are printed on a computer printout which contains a profile of the 20 need dimensions and a list of occupations showing how the individual's profile matches the needs of persons in these occupations.

4. Materials Required

   A. Test booklet

   B. Answer sheet

   C. Pencils

   D. Administration manual

   E. Two volumes of Occupational Reinforcer Patterns

5. Appropriate Groups - The MIQ is appropriate for any individual having at least a fifth grade reading level. Research has indicated that the MIQ is acceptable for use with minority and low socioeconomic status groups. Several MIQ users have utilized tape recorded administrations of the MIQ with individuals possessing low reading skills.

6. Technical Considerations

   A. Norm Groups - Norms were developed on a total of 5,358 individuals. The only comparison data needed for appropriate use of the MIQ consists of profiles of Occupational Reinforcer Patterns for each of the 148 occupations. This comparison data were collected from ratings of supervisors of incumbents within each occupation. Sample characteristics are fairly well defined.

   B. Reliability - Median internal consistency reliability coefficients for the 20 MIQ scales in several subject groups were generally in the .80's. Test-retest coefficients for the 20 MIQ scale scores ranged from a high of .89 for an immediate test-retest interval to a low of .48 for a six month test-retest interval. For MIQ profiles, however, median stability coefficients were in the .80's. Scale intercorrelations ranged from .05 to .77.

   C. Validity - Validation of the 1967 MIQ form consists of content validity studies, group difference studies, and concurrent validity studies. Reasonably good evidence of the validity of the MIQ was obtained using each of these methods. A summary of the results would be too detailed for this report.

7. APA Level - B

8. Sources of Information - MMY 8:1050; MMY 7:1063; Bolton, 236-237.

9. Available From - Vocational Psychology Research

10. Comments - The MIQ is a very appropriate tool for vocational evaluation. It is recommended that MIQ results be considered in conjunction with information concerning the abilities and interests of a client. The data presented in the MIQ can be very useful in helping a person clarify his value system as it relates to work.

OHIO VOCATIONAL INTEREST SURVEY (OVIS)

(1970)

1. Purpose - The OVIS was developed to counsel high school students on their future vocational plans.

2. Final Score - The third edition Dictionary of Occupational Titles (DOT) Data-People-Things organization was used as a model for development. The 24 OVIS interest scales represent combinations of the D-P-T classification. The 24 scales are: (1) manual work; (2) machine work; (3) personal services; (4) caring for people or animals; (5) clerical work; (6) inspecting and testing; (7) crafts and precise operations; (8) customer services; (9) nursing and related technical services; (10) skilled personal services; (11) training; (12) literary; (13) numerical; (14) appraisal; (15) agriculture; (16) applied technology; (17) promotion and communication; (18) management and supervision; (19) artistic; (20) sales representative; (21) music; (22) entertainment and performing arts, (23) teaching, counseling, and social work; and (24) medical. Final scores are presented on a computerized profile chart in scale scores, percentiles, and stanines. Summary reports are available for classes, schools, or special groups.

3. Description

   A. Administration - This pencil-and-paper test is administered to groups in between 60 and 90 minutes. The survey is not timed. Examinees record their responses on a separate answer sheet. The manual permits the examiner to read aloud the 27 items on the first part of the survey.

   B. Content - The OVIS is divided into two parts: (1) The Student Information Questionnaire asks the examinee for his stated interest, present educational status, and future career plans; (2) The Interest Inventory part has 280 items containing brief descriptions of activities. Each item requires the examinee to mark a five point scale form from "likes very much" to "dislikes very much." The reading level is estimated by the reviewer to be fifth grade.

   C. Scoring - The OVIS is machine scored only. Profile charts are returned for each examinee which give the results and explain the scales.

4. Materials Required

   A. Test booklet

   B. Answer sheet

   C. Pencils

   D. Administration manual

   E. Manual for interpretation

5. Appropriate Groups - The instrument was developed for use with high school students and appears to be most appropriate for use with young people in a school situation. Indeed, the first part of the OVIS would be almost meaningless for non-school persons. Because the OVIS is based on the DOT, it can be used for all job areas, thus, giving the widest possible occupational coverage. This broad occupational coverage makes it useful for persons with a wide range of abilities and occupational preferences. Because the items are written at a fairly low level, people with limited literacy skills should be able to complete the OVIS.

6. Technical Considerations

    A. Norm Groups - The OVIS was normed on over 45,000 students in grades 8 through 12 in ten states. Separate norms are used for each grade and each sex. Sample characteristics are briefly described.

    B. Reliability - Test-retest reliability coefficients (two week interval) were computed on samples of eighth and tenth grade students. These ranged from .73 to .90. These are not exceptionally high considering the short period of time between testing sessions.

    C. Validity - The manual presents no evidence of validity and bases its discussion of validity on the DOT.

7. APA Level - B

8. Sources of Information - MMY 8:1016; MMY 7:1029.

9. Available From - The Psychological Corporation

10. Comments - The use of the DOT as a model, the coverage of all occupational areas, the type of item, and the emphasis upon counseling the examinee on the results, make the OVIS a potentially useful device for vocational evaluators, especially those working with youth in an educational setting. The major problems are the lack of research, the requirement for computer scoring, and the content of the first part of the OVIS that is directly related to school attendance.

STRONG-CAMPBELL INTEREST INVENTORY (SCII)

Form T325 (1977)

1. Purpose - The SCII is the latest version of the Strong Vocational Interest Blank (SVIB) and the first one that combines the separate men's form (T399; T399R) and the women's form (W) into a single form for both sexes. Like the earlier SVIB, the SCII is intended to help guide persons into areas where they are likely to find the greatest job satisfaction. The manual states that the major use of the SCII is for vocational counseling.

2. Final Score - The following types of scores are given: (1) six general occupational themes (e.g., realistic, enterprising, and social); (2) 23 basic interest scales (e.g., nature, social service, and sales); (3) 124 occupational scales (e.g., occupational therapist, biologist, and lawyer); and (4) nine administrative and special indexes (e.g., school subjects, activities and amusements). Final scores are presented on a printout with standard scores and percentile bands, which indicate the percent of agreement between the client and the responses of the persons composing the scoring groups on which each particular scale was based.

3. Description

   A. Administration - SCII can be administered individually or to groups in about 30 minutes. The test is untimed and requires a sixth grade reading level. The items are printed on the answer sheet; no separate test booklet is required.

   B. Content - The SCII contains 325 items divided into seven sections: (1) occupations - 131 items; (2) school subjects - 36 items; (3) activities - 51 items; (4) amusements - 39 items; (5) types of people - 24 items; (6) preference between two activities - 30 items; and (7) characteristics - 14 items. For each item the respondent circles either "Like," "Dislike," or "Indifferent."

   C. Scoring - Machine scoring is the only method available. A computer printed profile of all the scores is returned.

4. Materials Required

   A. Answer sheet

   B. Pencils

   C. Manual

5. Appropriate Groups - The SCII was developed mostly on college samples and most of the occupations covered in the inventory are of a professional or semiprofessional nature. The item content requires a familiarity with many occupations and some advanced school subjects. For these two reasons, the SCII should be used only with persons who have the potential for advanced training. Because a person's interests often do

not solidify enough to be accurately measured prior to age 17, the manual contains a warning against using this test with persons under that age.

6. Technical Considerations

   A. Norm Groups - The Strong inventories have used a separate occupational group for each scale and then compare the results to men-in-general and women-in-general groups. The manual devotes considerable length to the problems of sex bias and differences in scores. All samples are clearly described. The items in the SCII are taken from the SVIB and, thus, the tremendous amount of data collected for the SVIB can apply to the newer SCII.

   B. Reliability - The manual presents test-retest correlations for the basic interest scales and the occupational scales for two weeks, 30 days, and three years. Median correlations were mostly in the low .90's. The inventory is designed to measure stable interest and the wealth of data on reliability suggests that they have accomplished this.

   C. Validity - Some validity data are presented in the manual for the SCII. However, "a substantial body of such data is available for the earlier SVIB . . .; because the new SCII scales are based directly on these earlier scales, these data are relevant for them also." Thus, the SCII is one of the most thoroughly researched and validated instruments available.

7. APA Level - B

8. Sources of Information - SCII: MMY 8:1023; Anastasi, 529-536; Bolton, 140-142. SVIB - Men's form: MMY 7:1036; MMY 6:1070; MMY 4:747; Women's form: MMY 7:1035; MMY 3:649.

9. Available From - Stanford University Press

10. Comments - The SCII (and the SVIB) have been used for years for counseling persons on vocational choices. The Strong has stood the tests of time and research and has become so widely accepted that it is the standard by which other interest inventories are judged. However, its use in an evaluation setting is hindered by its college orientation and item content requiring a knowledge of specific jobs and school subjects.

WIDE RANGE INTEREST-OPINION TEST (WRIOT)

(1972)

1. Purpose - The WRIOT "was designed to cover as many areas and levels of human activity as possible." This interest inventory contains items that represent jobs ranging from unskilled through the professional level.

2. Final Score - There are two types of scores: (1) 18 clusters of occupational interests (art, literature, music, drama, sales, management, office work, personal service, protective service, social service, social science, biological science, physical science, number, mechanics, machine operation, outdoor, and athletics) and (2) seven vocational attitudes (sedentariness, risk, ambition, chosen skill level, activity by sex, agreement, and interest spread). Scores are reported in T scores which are plotted on a profile form.

3. Description

   A. Administration - The untimed inventory can be given either individually or in groups. The estimated individual administration time is 40 minutes; 50 to 60 minutes for groups. The WRIOT items are illustrations which are presented in a spiral-bound booklet. In normal use the examinee responds through the use of a separate sheet. The manual contains special instructions for individual administration to "severely mentally or physically disabled persons." During individual administration, the client identifies the item and the evaluator records the response on the answer sheet.

   B. Content - The device consists of 450 line illustrations arranged in 150 triad combinations; from each triad the examinee picks the most liked and the least liked picture. The illustrations show men and women of all ethnic backgrounds engaging in a wide variety of job tasks from unskilled to professional. In most of the drawings the work activity is obvious.

   C. Scoring - The optical scanning answer sheets can be sent to the publisher for machine scoring or scored by hand. Twenty-four stencils are used in hand scoring. According to the manual, scoring and profiling can be done in about 20 minutes. Raw scores are converted into T scores using the appropriate norms table.

4. Materials Required

   A. Test booklet

   B. Answer sheet

   C. Pencils

   D. Administration manual

   E. Scoring stencils

   F. Report form

5. Appropriate Groups - The manual contains some specific instructions and a few hints on the application of the WRIOT to physically handicapped, mentally retarded, and disadvantaged persons. Some of the development was done with persons in sheltered employment, and the manual contains many examples of the use of the test in vocational rehabilitation. While the illustrations are generally easy to understand, the manual warns that "disadvantaged and sheltered people seem to be unfamiliar with some of the illustrated situations and place their own values on what they observe."

6. Technical Considerations

   A. Norm Groups - Separate norms are available by sex for the following groups: (1) adults, (2) grades 10 and 11, and (3) grade 8. The groups are fairly well described. However, the sample sizes range between 223 and 551 and all samples were taken from Delaware cities. There are no national samples and sampling techniques are not described.

   B. Reliability - The manual section on reliability opens with a poorly reasoned blast at those who require reliability coefficients and then goes on to give fairly high split-half reliabilities (mostly in the low .90's) for the 25 scales. Most interest inventories use test-retest measures of reliability.

   C. Validity - The only validity information presented are correlations between the WRIOT and the old Geist Picture Interest Inventory. Given the size of the groups and the faults of the Geist, these results are almost meaningless. Rather than presenting data, the manual takes the approach that the WRIOT is valid bacause its developers say it is valid.

7. APA Level - B

8. Sources of Information - MMY 8:1029.

9. Available From - Guidance Associates of Delaware, Inc.

10. Comments - The advantages of the WRIOT are (1) the use of a picture format and activities that cover the entire occupational spectrum and (2) the fact that the test can be used with persons having many handicapping conditions. The WRIOT successfully avoids the faulty reasoning that picture interest inventories are for persons who cannot read, therefore, these people are only interested in a few low level jobs. The disadvantages of the inventory are mostly technical: (1) the lack of an adequate norm base, (2) the weakness of the reliability data, and (3) the almost total disregard of the test's validity. In conclusion, this inventory is useful for a wide range of persons, but the results must be treated with extreme caution.

CRAWFORD SMALL PARTS DEXTERITY TEST (CRAWFORD)

(1956)

1. Purpose - This is a performance test designed "to measure fine eye-hand coordination."

2. Final Score - Percentile scores comparing the examinee's results with several norm groups are given for Part I (pins and collars) and Part II (screws).

3. Description

   A. Administration - This individually administered apparatus test takes about 15 minutes to complete. A work table and chair are required. The examinee sits during the test.

   B. Content - A board containing 42 holes each on the left and right bottom portions and three bins for pins, collars, and screws across the top portion is used. Part I requires the examinee to use tweezers to pick up one pin and place it in a hole on the board. He next uses the tweezers to fit a collar over each pin. After five pins and collars are assembled for practice, the examinee completes 36 pins and collars. In Part II the examinee uses a small screwdriver to screw 30 screws through a plate. Five screws are used for practice.

   C. Scoring - The amount of time required to complete Parts I and II is recorded for each part. The two time scores are compared to the appropriate norm tables.

4. Materials Required

   A. Apparatus board (including: pins, collars, screws, tweezers, and screwdriver)

   B. Manual

   C. Stopwatch

   D. Scoring sheets

5. Appropriate Groups - Because no reading is required, this test may be used with persons who are illiterate. Upper extremity handicaps could prevent the successful use of this test. The Crawford appears most useful in assessing persons for jobs involving the use of small tools and rapid, repeated movements.

6. Technical Considerations

   A. Norm Groups - Percentile norms based on the time to completion are available on the following male groups: unselected applicants, appliance factory applicants, two veterans groups, and two high school groups. Female norm groups are: assembly job applicants, factory applicants, hourly employees, and employed assemblers. All sample sizes are at least 100 and most are over 175. Samples are inadequately described.

B. Reliability - Split-half reliabilities for Parts I and II are reported as being between .80 and .95. However, since the Spearman-Brown formula was used, these are overestimates. No test-retest correlations are provided--these would have been a more appropriate measure of reliability.

C. Validity - The manual contains summaries of several validation studies using wages and supervisors' ratings as criteria. The test is also related to other dexterity measures.

7. APA Level - A

8. Sources of Information - MMY 5:871; MMY 4:752; MMY 3:667; Anastasi, 444; Bolton, 260; Guion, 292-293.

9. Available From - The Psychological Corporation

10. Comments - The Crawford could be used as one method of assessing a client's finger dexterity. Since it is recommended that the test be given individually, the evaluator could use this as a basis for some detailed behavior observations.

HAND-TOOL DEXTERITY TEST (BENNETT)

1965 Revision (1965)

1. Purpose - The purpose is to "provide a measure of proficiency in using ordinary mechanics' hand-tools . . . The ability measured by this test is a combination of aptitude and of achievement based on past experience in handling tools."

2. Final Score - A single final score is compared with one of several groups; this final score given in percentiles is based upon the time to complete the task.

3. Description

   A. Administration - This individually administered apparatus test is untimed; most examinees complete it in between five and 20 minutes. The apparatus frame should be bolted to a sturdy work table 34 inches high. The examinee stands during this test.

   B. Content - Three different sizes (four each) of nuts, bolts, and washers are removed from one side of a hardwood frame with the aid of three wrenches and one screwdriver. The nuts and bolts are then fastened and tightened through the holes on the other side of the frame. The test administrator first reads the instructions and then times the examinee with a stopwatch.

   C. Scoring - One time score in minutes and seconds to completion is obtained and this is compared with norm tables. Comparison with norm tables takes less than one minute.

4. Materials Required

   A. Apparatus (consisting of: frame, screwdriver, wrenches, nuts, bolts, and washers)

   B. Manual

   C. Stopwatch

   D. Scoring sheets

5. Appropriate Groups - Because no reading ability is required and because the manual permits the administrator "to supplement the directions in any reasonable way to improve the examinee's understanding of the task," literacy and difficulty in understanding instructions should not be problems. The test does require full use of hands and arms and, therefore, may not be appropriate for people with an upper extremity handicap. The "face" validity of the test may appeal to the disadvantaged and clients with low motivational levels.

6. Technical Considerations

   A. Norm Groups - Percentile norms based on the time to completion are given in the manual for the following groups: male job applicants in a southern plant, male adults at a vocational guidance center, airline engine mechanics, apprentice welders in a steel company, electrical maintenance workers, employees and applicants in a manufacturing company, boys at a vocational high school, and high school dropouts in a metropolitan center. The composition of the eight norm groups are not adequately described in the manual. The mean age, job experience, minority group status, and other important descriptive information is not presented. Also the norms for four of the eight groups are based on sample sizes of less than 200 subjects; these should be used with extreme caution.

   B. Reliability - The manual reports two test-retest studies which produced reliability coefficients of .91 and .81, considered moderately high for a performance test.

   C. Validity - Two types of validity data are presented: (1) correlations with foremen's ratings and (2) correlations with other tests. Foremen's ratings were between .14 and .51; other tests between .11 and .42.

7. APA Level - A

8. Sources of Information - MMY 7:1044; MMY 3:659.

9. Available From - The Psychological Corporation

10. Comments - The high face validity, flexibility of the instructions and the practical nature of the test are appealing in a vocational evaluation situation. The norm groups, however, need to be described in greater detail and caution should be used in interpretation of results based on some norm groups.

PURDUE PEGBOARD

(1948)

1. Purpose - This dexterity test is designed to aid in the selection of employees for industrial jobs requiring manual dexterity. It measures dexterity for two types of activity: one involving gross movements of hands, fingers, and arms, and the other involving primarily what might be called "fingertip" dexterity.

2. Final Score - Five separate percentile scores are obtained: (1) right hand, (2) left hand, (3) both hands, (4) right plus left plus both hands, and (5) assembly.

3. Description

   A. Administration - The apparatus test can be administered to seated groups of up to ten persons in about ten minutes. Besides the pegboard, no special conditions or materials are necessary. There are four sub-tests: (1) right hand only, (2) left hand only, (3) both hands simultaneously, and (4) assembly. No tools are used.

   B. Content - The Pegboard includes pins, collars, and washers which are located in four cups at the top of the board. Each sub-test involves a separate task. The right hand test involves placing pins into holes on the board for a 30 second period. Left hand involves the same process, but with the opposite hand. Both hands involves placing pins as fast as possible into holes simultaneously with both hands. The right plus left plus both hands is obtained by adding the above three scores together. The assembly task consists of assembling pins, collars and washers on the board for a speeded time of one minute.

   C. Scoring - The scores are the number of pins placed in the board within the 30 second time limit for the right, left, and both hands, and the number of assemblies completed in one minute. Results are compared to norms tables.

4. Materials Required

   A. Pegboard (including pins, collars, and washers)

   B. Manual

   C. Recording sheets

   D. Stopwatch

5. Appropriate Groups - Because reading ability is not required and because of the demonstrations, literacy should not be considered a barrier for administration. The test is administered to seated examinees and may be appropriate for some wheelchair clients. The test requires full use of hands and arms and, therefore, may not be appropriate for clients having upper extremity handicaps.

6. Technical Considerations

   A. Norm Groups - Percentile norms are given for eight groups of male and female industrial workers and two general groups composed of applicants and college students. The manual describes most of the characteristics of these groups except their job experience, minority group composition, and date of norming.

   B. Reliability - The test-retest reliability correlations are between .66 and .79. These are low for a standardized test and, therefore, the test should be used with caution, especially when recommending employment based on its results.

   C. Validity - Correlations with production records, supervisor's ratings, and a job sample range between .09 and .61.

7. APA Level - A

8. Sources of Information - MMY 6:1081; MMY 5:873; MMY 4:751; MMY 3:666; Anastasi, 444; Guion, 291-292; Horrocks & Schoonover, 379-380; Super & Crites, 213-217.

9. Available From - Science Research Associates

10. Comments - Because of low reliability and low validity coefficients presented in the manual, it appears that the test stands mostly on face validity. It is suggested that the test not be used alone to select people for assembly jobs; the test is best used as part of a battery of tests and work samples.

STROMBERG DEXTERITY TEST (STROMBERG or SDT)

(1951)

1. Purpose - The Stromberg "was developed as an aid in choosing workers for jobs which require speed and accuracy of arm and hand movement."

2. Final Score - Percentile scores compare the examinee's results with several norm groups.

3. Description

   A. Administration - This individually administered apparatus test can be given and scored in about ten minutes. The examinee stands at a 30 inch table to take the test.

   B. Content - The Stromberg contains 54 round red, blue, and yellow discs with both sides painted and a durable board containing 54 holes on one side of the hinges. The other side consists of a flat surface only. (The board folds in half for storage.) The SDT uses four trials. On the first trial the examinee transfers the discs according to a set pattern from the form board to the open board. On the second trial the discs are moved back to the form board holes. Trials three and four are identical to trials one and two. All movement is with one hand.

   C. Scoring - The examinee receives four trials; trials one and two are practice and are not scored. The number of seconds needed to complete trials three and four are added to obtain a single final score. This final score is compared to the various norm tables.

4. Materials Required

   A. Form board

   B. Manual

   C. Stopwatch

   D. Recording sheets

5. Appropriate Groups - The Stromberg can be used for most clients who can stand, rotate their torso, use one upper extremity, and see. The SDT is intended for selecting persons in jobs where gross, rapid hand and arm movements are required.

6. Technical Considerations

   A. Norm Groups - Seven norm groups are available, ranging from trade school students through male and female applicants to male and female workers. No descriptive data is given about the characteristics of the norm groups; not even their sizes are presented.

B. Reliability - The manual lists two studies using the Spearman-Brown formula yielding correlations of .84 and .87. Because this formula overestimates the reliability, the "true" reliability of the Stromberg is somewhat lower than the reported coefficients.

C. Validity - Validity data is of two types: (1) "good" and "poor" workers, as determined by their wages, were compared on basic scores on the Stromberg; "good" workers had higher test scores, (2) correlations with two other dexterity tests yielded correlations averaging in the low .30's.

7. APA Level - A

8. Sources of Information - MMY 4:755

9. Available From - The Psychological Corporation

10. Comments - The Stromberg is of limited use in selection because of its inadequate norms and low reliability. However, it has potential use in assessing client dexterity if supplemented with other test and work sample results.

BENNETT MECHANICAL COMPREHENSION TEST (BENNETT OR BMCT)

Forms S and T (1969)

1. Purpose - The BMCT measures "the ability to perceive and understand the relationship of physical forces and mechanical elements in practical situations." It is intended to be used with job applicants, high school students, and candidates for engineering schools.

2. Final Score - The examinee receives a single mechanical comprehension score in percentile form.

3. Description

   A. Administration - BMCT is a group administered pencil-and-paper test. Although it has a 30 minute time limit, the test is essentially a power test. A separate answer sheet is used.

   B. Content - Most of the 68 items contain two illustrations and a written question asking which is easier to turn, which will move in a certain direction, etc. The items deal with gears, hydraulics, pulley systems, structures, levers, center of gravity, etc.

   C. Scoring - Scoring is by hand or machine. Hand scoring takes about three minutes and scores are compared to one or more of the norm groups listed in the manual.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Answer sheet

   D. Manual

   E. Stopwatch

   F. Scoring stencil

5. Appropriate Groups - The illustrations make the test interesting and provide a high "face" validity. Although the format, printing, etc., of the test booklets is of a high quality, the examinee will need good visual acuity for many of the illustrations, especially those showing gears and pulleys. The BMCT does not appear to be too heavily culturally biased.

6. Technical Considerations

   A. Norm Groups - The test has norms based on three general groups of males: (1) industrial applicants (five separate norm groups), (2) industrial employees, and (3) students (four separate norm groups). Sample sizes vary between 85 and 906, with most of the groups having about 100 subjects. Groups are described in sufficient detail.

B. Reliability - Even-odd reliability coefficients are given for eight different groups; these range between .81 and .93. Because the Spearman-Brown formula was used, these are overestimates.

C. Validity - The manual presents numerous studies showing correlations between the BMCT and criteria such as course grades, job ratings, and training success. Of additional interest are the correlations between this test and other mechanical and ability tests. Because most of these are with earlier forms of the test, their usefulness is questionable even though most of the items in the present forms were taken from the older forms.

7. APA Level - A

8. Sources of Information - MMY 7:1049; MMY 6:1094; MMY 5:889; MMY 4:776; MMY 3:683; Anastasi, 448-449; Super & Crites, 242-256.

9. Available From - The Psychological Corporation

10. Comments - This test has had a long history of use in selection and placement of individuals on jobs where mechanical comprehension is required. Because the BMCT has a high face validity and the illustrations are interesting (especially for male clients), it may be useful for people who are difficult to test. Because of the size of some of the norm groups, caution is necessary when interpreting results. The manual also contains a special section on mechanical ability in women and how the BMCT attempts to deal with this ability in women in a valid and yet practical way.

THE REVISED MINNESOTA PAPER FORM BOARD TEST

Series AA and BB; MA and MB (1969)

1. Purpose - The form board has two major purposes: (1) the original purpose was in selection and placement in jobs requiring a "mechanical orientation" and (2) the test also measures "spatial imagery," which correlates with general intelligence, thus, providing a nonverbal estimate of intellectual functioning.

2. Final Score - A single, final score is given in percentiles.

3. Description

   A. Administration - The Form Board is a group pencil-and-paper administered test. Administration time is 20 minutes. For Series AA and BB, the examinee records his answer on the test itself; Series MA and MB require a separate answer sheet.

   B. Content - "The items consist of 64 two-dimensional diagrams cut into separate parts. For each diagram there are five figures with lines indicating the different shapes out of which they are made." The examinee chooses one figure "composed of the exact parts that are shown in the original diagram."

   C. Scoring - Series MA and MB are machine scored; Series AA and BB are hand scored. Hand scoring takes about five minutes and the final score is compared to the appropriate norm group.

4. Materials Required

   A. Test booklet

   B. Answer sheet (Series MA and MB only)

   C. Pencils

   D. Manual

   E. Scoring stencils

   F. Stopwatch

5. Appropriate Groups - Because the test does not require reading and the easy-to-understand format of Series AA and BB, the Form Board appears to be appropriate for persons with low literacy and test-taking skills. The items are small in size requiring good visual acuity. The abstract nature of the test items mean that the evaluator will want to make certain that his/her clients understand the reason for the test.

6. Technical Considerations

   A. Norm Groups - The device gives educational (grades 10 through 12) and industrial norms (applicants, employed workers, and military) for each series. The samples are carefully described and the groups large enough to provide meaningful standards of comparison.

   B. Reliability - The manual contains alternate-form and test-retest reliability coefficients. Alternate form correlations range from .71 to .78; test-retest from .79 to .90. Both types of reliability are quite high.

   C. Validity - Validity is of two types: (1) comparison of the scores of different groups and (2) correlations with other tests. The Form Board is capable of distinguishing between success and failure and is related to other tests measuring similar abilities.

7. APA Level - A

8. Sources of Information - MMY 7:1056; MMY 6:1092; MMY 5:884; MMY 4:763; Anastasi, 447-448; Guion, 246; Horrocks & Schoonover, 371-375; Super & Crites, 622.

9. Available From - The Psychological Corporation

10. Comments - This technically adequate test has a long history of successful research and practical use. The authors of the test, realizing that the Form Board is only one part of a valid assessment of mechanical aptitude and prediction of job success, suggest that the Form Board be used as part of a battery.

SRA MECHANICAL APTITUDES

(1950)

1. Purpose - This test is intended to measure three aspects of mechanical ability (see below). The concept behind the SRA Mechanical Aptitudes is that "no single test consisting of items which measure only one primary component of mechanical aptitude can provide a satisfactory index of the ability to learn mechanical skills."

2. Final Score - The test yields four final scores: (1) mechanical knowledge, (2) space relations, (3) shop arithmetic, and (4) a total score based on the sum of the previous three. Scores are presented in percentiles.

3. Description

   A. Administration - This group administered multiple-choice pencil-and-paper test can be given in about 40 minutes. While the three parts of the test are timed, the test does not appear to be highly speeded. A separate self-scoring carbon-centered answer sheet is used.

   B. Content - The test contains three sections: (1) mechanical knowledge - 45 pictures of commonly used tools and implements are identified and/or their use is given, (2) space relations - 40 figures cut into two or three pieces which are mentally formed into a whole, and (3) shop arithmetic - 124 problems, most of which are based on drawings. The manual contains no estimated reading level, but the reviewer estimates that a fifth grade level is needed.

   C. Scoring - The test is hand scored using a carbon-centered answer sheet. Raw scores are compared to norms and plotted on a separate profile sheet. Scoring time should be less than three minutes.

4. Materials Required

   A. Test booklet

   B. Pencils

   C. Answer sheet

   D. Scratch paper

   E. Administration booklet

   F. Stopwatch

   G. Profile sheet

5. Appropriate Groups - The SRA Mechanical Aptitudes is designed to assess the mechanical ability of persons planning careers or applying for jobs in mechanical areas. The test's limitations are: (1) the reading level

and (2) the visual acuity required to see the tools, shapes, and the diagrams relating to the arithmetic section. While all of these are simple black-on-white line drawings, some may be difficult to visualize. Because the answer sheet must be placed on the right side of the test booklet, left-handed examinees will often need to cover the items while recording answers.

6. Technical Considerations

   A. Norm Groups - Separate norms are given by sex and grade for grades 9 through 12. Norms are also given on 650 male trainees. Although all norm groups are of adequate size, the sample characteristics are not described in enough detail to permit an accurate judgment of their relevance.

   B. Reliability - Internal consistency estimates using the KR-21 formula are given for the school boys groups on each score. They range from .55 to .83, which are acceptable.

   C. Validity - No validity data are contained in the manual.

7. APA Level - A

8. Sources of Information - MMY 4:764

9. Available From - Science Research Associates

10. Comments - The major feature of this test is that it combines three aspects of "mechanical aptitude" into one test with a high "face" validity. While the concept is a good one, the manual is over 25 years old and needs to be updated before it can be expected to meet current standards.

SOURCES OF INFORMATION ABOUT TESTING

Books on Testing and Statistics

This section provides a listing of reference materials about testing and statistics. Most titles were selected because these books are readily available; many being used as textbooks in college-level courses. These two lists are not intended to be inclusive; a complete list of books on testing and statistics would contain several hundred titles. It is hoped that the interested reader will consult these materials as needed.

Testing books:

1. Anastasi, A., Psychological testing (fourth edition). New York: Macmillan Publishing Co., Inc., 1976. (750 pages)

"The primary goal of this text is . . . to contribute toward the proper evaluation of psychological tests and the correct interpretation and use of test results. This goal calls for several kinds of information: (1) an understanding of the major principles of test construction, (2) psychological knowledge about the behavior being assessed, (3) sensitivity to the social and ethical implications of test use, and (4) broad familiarity with the types of available instruments and the sources of information about tests."

2. Barnette, W. L., Jr. (Ed.), Readings in psychological tests and measurements (rev. ed.). Homewood, Illinois: Dorsey Press, 1968. (393 pages)

This book of readings is aimed at undergraduate psychology majors with a limited knowledge of advanced statistics. Most of the articles have been edited and shortened. Some of the topics covered are test administration problems, response sets, validity, and public policy on testing.

3. Bolton, B. (Ed.), Handbook of measurement and evaluation in rehabilitation. Baltimore: University Park Press, 1976. (362 pages)

This recent publication is a collection of 17 papers designed to "summarize the current status of . . . psychological measurement principles and practices as they are applied in the evaluation of disabled clients." This work provides for a reference for professionals who must interpret psychological evaluation reports. In addition to three chapters on norms, reliability and validity, the publication contains several chapters reviewing specific tests. There are also chapters reviewing the uses of testing in rehabilitation.

4. Cronbach, L. J., Essentials of psychological testing (third edition). New York: Harper and Row, 1970. (752 pages)

This book is probably the best known and widely used of all college texts. It is designed to be a comprehensive introduction to the uses of testing, types of tests, administration, scoring, and validation. Other parts of the book provide descriptions and reviews of a wide variety of tests.

5. Ghiselli, E. E., Theory of psychological measurement. New York: McGraw-Hill, 1964. (408 pages)

This book does not discuss individual tests, rather provides the logical and mathematical reasons behind test construction and subsequent test use. Designed to cover the basic problems of psychological measurement, concepts are clearly explained with simple examples and all formulas are given in detail.

6. Guion, R. M., Personnel testing. New York: McGraw-Hill, 1965. (585 pages)

"This book is principally concerned with employment problems and with the implications of psychological testing methods for these problems." Here the "applications of the techniques and principles of testing are applied to personnel selection and related employment problems." Because it emphasizes the unsolved problems in the area of selection and requires a prior knowledge of testing, this publication is for persons interested in an in-depth study of selection problems.

7. Horrocks, J. E., & Schoonover, T. L., Measurement for teachers. Columbus, Ohio: Charles E. Merrill, 1968. (645 pages)

"This book has been written specifically for teachers, counselors and others who deal with children in a school setting." The book is strong on the measurement of achievement (i.e., reading, mathematics, and language arts), but also contains sections on intelligence and interests. The authors emphasize that this is a book designed for a course in measurement and not a course in statistics.

8. Jackson, D. N., & Messick, S., (Eds.), Problems in human assessment. New York: McGraw-Hill Book Co., 1967. (873 pages)

This large collection of articles is aimed at the advanced undergraduate or beginning graduate student. The 74 contributions cover the logic of assessment, assessment methodology, selection, intellectual abilities, personality, attitudes, and ethics. The publication contains many of the "classic" articles in the area of assessment.

9. Lyman, H. B., Test scores and what they mean. Englewood Cliffs, New Jersey: Prentice-Hall, 1963. (223 pages)

This book deals mainly with the meaning of test scores and is "intended to meet the needs of test users with limited training in testing . . ." Real-life examples are used to illustrate the major concepts. A unique factor is an original classification of the types of test scores.

10. Maloney, M. P., & Ward, M. P., Psychological assessment: A conceptual approach. New York: Oxford University Press, 1976. (422 pages)

The authors' attempt to formulate a conceptual model for the use of testing, interviewing, and case histories in the assessment of persons with mental and emotional disabilities. Sections on report writing are also included. The emphasis is upon the process of assessment and not on specific tests. The theme of the publication appears to be: "Most criticisms are not leveled at tests, per se, but at the use of these tools in practice."

11. Noll, V. H., & Scannell, D. P., Introduction to educational measurement (third edition). New York: Houghton Mifflin Co., 1972. (592 pages)

". . . Our purposes are to provide an orientation to the field on measurement and evaluation, a foundation in measurement theory and elementary statistical methods, an acquaintance with published standardized tests and sources of information about them, basic understanding and skill in constructing tests for local use, and instruction in the interpretation and application of the results of measurement."

12. Super, D. E., & Crites, J. O., Appraising vocational fitness (rev. ed.) New York: Harper and Brothers, 1962.

This somewhat dated publication presents a traditional view of testing with emphasis upon its uses in prediction of training and job success. The book includes critical reviews of many of the tests commonly used in education and industry.

13. Thorndike, R. L., Personnel selection. New York: John Wiley & Sons, 1949. (358 pages)

"This book is made up of two rather distinct parts. The first eight chapters deal with the technical problems involved in developing a personnel testing program and in appraising its effectiveness. The last three deal with administrative problems of maintaining an effective, smooth-running program with good public acceptance." This now classic book on industrial testing covers everything from job analysis to the operation of a testing program.

14. Thorndike, R. L., & Hagen, E., Measurement and evaluation in psychology and education (third edition). New York: John Wiley and Sons, 1969. (705 pages)

This intermediate level text covers almost all phases of testing in detail and gives the reader the reasons behind many common testing practices. In addition to containing detailed discussions of the basic measurement concepts, the book also provides critical reviews, as well as illustrative items, of many commonly used tests.

15. Tyler, L. E., Tests and measurements (second edition). Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971. (199 pages)

Intended for use as a beginning text in measurement, this short book attempts to cover basic statistics, test theory, as well as some special types and uses of tests.

Statistics Books - Introductory Level

1. Bruning, J., Computational handbook of statistics. Glenview, Illinois: Scott, Foresman, and Co., 1968.

2. Byrkit, D., Elements of statistics (second edition). New York: Van-Nostrand Co., 1975.

3. Gatkin, L., & Goldstein, L., Descriptive statistics: A programmed textbook (Volumes I & II). New York: John Wiley and Sons, Inc.

4. Hamburg, M., Basic statistics: A modern approach. New York: Harcourt, Jovanovich, Inc., 1974.

5. McCollough, C., & Vanalla, L., Introduction to description statistics and correlation: A program for self-instruction. New York: McGraw-Hill, 1965.

Statistics Books - Intermediate Level

1. Bloomers, P. J., Elementary statistical methods in psychology and education. Boston: Houghton Mifflin, 1960.

2. Gourevitch, V., Statistical methods: A problem-solving approach. Boston: Allyn and Bacon, 1965.

3. Guilford, J. P., Fundamental statistics in psychology and education (fourth edition). New York: McGraw-Hill, 1965.

4. Hays, W. L., Statistics for psychologists. New York: Holt, Rinehart, & Winston, 1963.

5. Runyon, R. P., & Haber, A., Fundamentals of behavioral statistics. Reading, Massachusetts: Addison-Wesley, 1967.

6. Walpole, R. E., Introduction to statistics. New York: MacMillan, 1968.

Mental Measurements Yearbooks

Perhaps the best single source of accurate information on tests are the Mental Measurements Yearbooks (MMY)*. Edited by Oscar K. Buros, the first MMY was published in 1938 and the eighth MMY in November 1978. Each one contains carefully worded descriptions of tests, critical reviews, and references to research studies using the test. Every MMY attempts to present the most recent information available about each test.

In addition to the MMY's, the publisher also has two other publications on tests. The first, Tests in Print II, contains a thorough listing and brief description of all tests that were published before 1974. Tests in Print II also contains many references for tests. The second, Vocational Tests and Reviews, is a composite of all reviews on vocational tests from the first to the seventh MMY.

The Casebook on Ethical Standards of Psychologists

Ethical problems related to test security and test interpretation are discussed and examples are given. The book also gives the purchasing requirements for tests. The sale of tests is restricted in accordance with principles given in the Ethical Standards*. Eligibility to purchase tests is determined on the basis of training and experience. Registration forms from many publishers are available upon request. The following classes are usually used:

---

*Published by: The Gryphon Press, 220 Montgomery Street, Highland Park, New Jersey 08904.

1. Schools, Colleges, and Governmental Agencies

   Orders received on official purchase forms or by officially signed letters will be filled promptly - orders may have to be countersigned by professor who assumes responsibility, if graduate student, etc.

2. Business and Industrial Firms

   Level A--Company purchase orders for tests commonly used for employment purposes will be filled promptly.

   Level B--Staff member of firm must have completed advanced level course in testing at University or its equivalent in training under qualified superintendent.

   Level C--Available to firms only for use by qualified psychologists, members of the American Psychological Association or person with Master's degree in psychology and appropriate training in field of personnel testing.

3. Consultants to business and industry, employment agencies, vocational counselors, and psychologists in private practice. Registration is required. Approval for test purchase is granted or withheld. No tests are sold for self-guidance, nor to any agency engaged in testing by mail.

Item seven of the Test Review Outline refers to the APA Level of the test. This is the restriction code of A, B, or C described above.

Standards for Educational and Psychological Tests and Manuals*

This publication contains technical standards which testing materials should meet. These standards cover the information that should be contained in test manuals, how reliability and validity should be reported, and how tests should be used. It is urged that the evaluator use this publication as a guide in judging the technical quality of testing materials.

---

*Published by: American Psychological Association, Inc., 1200 Seventeenth St. N.W., Washington, D.C. 20036.

ADDRESSES OF TEST PUBLISHERS

American Association on Mental Deficiency
5201 Connecticut Avenue, N.W.
Washington, D.C. 20015

American Guidance Service, Inc.
Publishers' Building
Circle Pines, Minnesota 55014

Bobbs-Merrill Co.
4300 West 62nd Street
Indianapolis, Indiana 46268

CTB/McGraw-Hill
Del Monte Research Park
Monterey, California 93940

Guidance Associates of Delaware, Inc.
1526 Gilpin Avenue
Wilmington, Delaware 19806

Houghton Mifflin Co.
110 Tremont Street
Boston, Massachusetts 02107

Institute for Personality and Ability Testing
1602 Coronado Drive
Champaign, Illinois 61820

The Psychological Corporation
757 Third Avenue
New York, New York 10017

Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611

Stanford University Press
Stanford, California 94305

Vocational Psychology Research
Elliott Hall
University of Minnesota
Minneapolis, Minnesota 55455

REFERENCES

Bauman, M. K., Psychological evaluation of the blind client. In B. Bolton (Ed.), Handbook of measurement and evaluation in rehabilitation. Baltimore: University Park Press, 1976, 249-268.

Botterbusch, K. F., Test use in vocational evaluation. Vocational Evaluation and Work Adjustment Bulletin, 1974, 7(2), 6-10.

Botterbusch, K. F., The use of psychological tests with individuals who are severely disabled. Menomonie, Wisconsin: Materials Development Center, 1976.

Botterbusch, K. F., & Sax, A. B., Some considerations for selection of a commercial vocational evaluation system. Vocational Evaluation and Work Adjustment Bulletin, 1977, 10(3), 32-37.

Coffey, D. D., Vocational evaluator competencies and their relative importance as perceived by practitioners and educators in vocational evaluation. Unpublished doctoral dissertation, Auburn University, 1978.

Commission on the Accreditation of Rehabilitation Facilities, Standards manual for rehabilitation facilities. Chicago: Author, 1978.

Dickson, M., Work sample evaluation of blind clients: Criteria for administration and development. Menomonie, Wisconsin: Materials Development Center, 1976.

Egerman, K., & Gilbert, J., The work evaluator. Journal of Rehabilitation, 1969, 35(3), 12-14.

Levine, E., Psychological evaluation of deaf clients. In B. Bolton (Ed.), Handbook of measurement and evaluation in rehabilitation. Baltimore: University Park Press, 1976, 269-292.

Pruitt, W. A., Task analysis of the vocational rehabilitation graduate major with an emphasis in work evaluation: A comparative study of two groups of work evaluators. Menomonie, Wisconsin: Department of Rehabilitation and Manpower Services, October, 1972.

Sankovsky, R., Characteristics of the VEWAA membership: 1971. Vocational Evaluation and Work Adjustment Bulletin, 1971, 4, 17-26.

Scholl, G., & Schnur, R., Measures of psychological, vocational, and educational functioning in the blind and visually handicapped. New York: American Foundation for the Blind, 1976.

Task Force No. 3, The evaluator. Vocational Evaluation and Work Adjustment Bulletin, 1975, 8(3), 121-141 (special issue, September, 1975).